

التعلم الآلي والتعلم العميق وعلم البيانات

ترجمة واعداد:

د. علاء طعيمة

بســمه تعالى

معجم مصطلحات

التعلم الآلي والتعلم العميق وعلم البيانات

ترجمة واعداد:

د. علاء طعيمة

مقدمة المؤلف

عند تعلم التعلم الآلي والتعلم العميق وعلم البيانات، فإنك تمر بالعديد من المصطلحات التي ليس من السهل تذكرها، ولكن إذا كان لديك معجم مصطلحات التعلم الآلي والتعلم العميق وعلم البيانات، فسيساعدك ذلك على تذكر معنى المصطلحات التي لا تتذكرها بسهولة. لذلك في هذه الكتاب، سوف أقدم لكم معجم مصطلحات التعلم الآلي والتعلم العميق وعلم البيانات الذي قمت بإنشائه لتسهيل تذكر معنى المصطلحات الصعبة في هذه المجالات. يحتوي هذا المعجم على أهم مصطلحات التعلم الآلي والتعلم العميق وعلم البيانات التي تحتاج إلى معرفتها عند تعلم التعلم الآلي والتعلم العميق وعلم البيانات. نأمل أن يساعدك معجم التعلم الآلي والتعلم العميق وعلم البيانات هذا في تذكر معاني المصطلحات التي لا يمكنك تذكرها بسهولة.

لقد حاولت قدر المستطاع ان اجمع المصطلحات الأكثر طرحاً مع الشرح المناسب والكافي، ومع هذا يبقى عملاً بشرياً يحتمل النقص، فاذا كان لديك أي ملاحظات حول هذا الكتاب، فلا تتردد بمراسلتنا عبر بريدنا الالكتروني alaa.taima@qu.edu.iq.

نأمل ان يساعد هذا الكتاب كل من يريد ان يدخل في مجالات التعلم الآلي والتعلم العميق وعلم البيانات ومساعدة القارئ العربي على تعلم هذا المجالات. اسأل الله التوفيق في هذا العمل لأثراء المحتوى العربي الذي يفتقر أشد الافتقار إلى محتوى جيد ورصين في مجال التعلم الآلي والتعلم العميق وعلم البيانات. ونرجو لك الاستمتاع مع الكتاب ولا تنسونا من صالح الدعاء.

د. علاء طعيمة

كلية علوم الحاسوب وتكنولوجيا المعلومات

جامعة القادسية

العراق

المحتويات


أهم جروبات علي تليجرام
باحثون
هنا سور الأزبكية
غواص في بحر الكتب
قناة مصر الثقافية و الفنية

A



A

A/B testing

اختبار A / B

طريقة إحصائية لمقارنة تقنيتين (أو أكثر) A و B. نموذجيًا، A هي تقنية موجودة، وB هي تقنية جديدة. لا يحدد اختبار A / B الأسلوب الأفضل فحسب، بل يحدد أيضًا ما إذا كان الاختلاف ذا دلالة إحصائية.

يقارن اختبار A / B عادةً مقياسًا (metric) واحدًا على طريقتين؛ على سبيل المثال، كيف تقارن دقة (accuracy) النموذج لتقنيتين؟ ومع ذلك، يمكن أن يقارن اختبار A / B أيضًا أي عدد محدود من المقاييس.

Accuracy

الدقة

عدد تنبؤات التصنيف الصحيحة مقسومًا على العدد الإجمالي للتنبؤات. هذا هو:

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{correct predictions} + \text{incorrect predictions}}$$

على سبيل المثال، النموذج الذي أجرى 40 تنبوًا صحيحًا و10 تنبؤات غير صحيحة سيكون له دقة:

$$\text{Accuracy} = \frac{40}{40 + 10} = 80\%$$

يوفر التصنيف الثنائي (Binary classification) أسماء محددة لفئات مختلفة من التنبؤات الصحيحة والتنبؤات غير الصحيحة. إذن، معادلة الدقة للتصنيف الثنائي هي كما يلي:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

حيث:

- TP هو عدد الإيجابيات الحقيقية (True Positives) (تنبؤات صحيحة).
- TN هو عدد السلبيات الحقيقية (True Negatives) (تنبؤات صحيحة).
- FP هو عدد الإيجابيات الخاطئة (False Positives) (تنبؤات غير صحيحة).
- FN هو عدد السلبيات الخاطئة (False Negatives) (تنبؤات غير صحيحة).

قارن الدقة (accuracy) مع الدقة (precision) والاسترجاع (recall).

على الرغم من كونه مقياسًا قيمًا لبعض المواقف، إلا أن الدقة (accuracy) مضللة للغاية بالنسبة للآخرين. والجدير بالذكر أن الدقة (accuracy) عادةً ما تكون مقياسًا ضعيفًا لتقييم نماذج

التصنيف التي تعالج مجموعات البيانات غير المتوازنةفي الفئة (class-imbalanced datasets).

على سبيل المثال، لنفترض أن الثلج يتساقط 25 يومًا فقطفي كل قرنفي مدينة شبه استوائية معينة. نظرًا لأن الأيام الخالية من الثلج (الفئة السلبية Negative class) فاق عدد الأيام التي بها تساقط للثلوج (الفئة الإيجابية positive class)، فإن مجموعة بيانات الثلج لهذه المدينة غير متوازنة طبقًا. تخيل نموذج تصنيف ثنائي من المفترض أن يتوقع إما تساقط ثلوج أو عدم وجود ثلوج كل يوم ولكنه ببساطة يتوقع "عدم وجود ثلوج" كل يوم. هذا النموذج دقيق للغاية ولكن ليس له قوة تنبؤية. يلخص الجدول التالي نتائج قرن من التنبؤات:

Category	Number
TP	0
TN	36500
FP	25
FN	0

لذلك فإن دقة هذا النموذج هي:

accuracy = (TP + TN) / (TP + TN + FP + FN)
accuracy = (0 + 36500) / (0 + 36500 + 25 + 0) = 0.9993 = 99.93%

على الرغم من أن الدقة البالغة 99.93٪ تبدو نسبة رائعة جداً، إلا أن النموذج ليس لديهفي الواقع قوة تنبؤية.

عادةً ما تكون الدقة (Precision) والاسترجاع (recall) مقاييس أكثر فائدة من الدقةفي تقييم النماذج المدربة على مجموعات البيانات غير المتوازنةفي الفئة.

Action

الاجراء

في التعلم المعزز (reinforcement learning)، الآلية التي من خلالها ينتقل الوكيل (agent) بين حالات (states) البيئة (environment). يختار الوكيل الإجراء باستخدام سياسة (policy).

Activation function

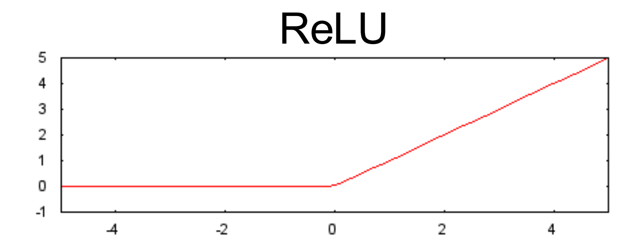
دالة التنشيط

دالة تمكن الشبكات العصبية (neural networks) من تعلم العلاقات غير الخطية (nonlinear) (المعقدة) بين الميزات والتسمية.

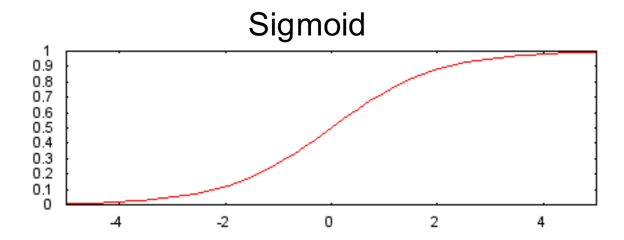
تتضمن دوال التنشيط الشائعة ما يلي:

- ReLU
- Sigmoid

مخططات دوال التنشيط ليست خطوطًا مستقيمة مفردة أبدًا. على سبيل المثال، تتكون دالة تنشيط ReLU من خطين مستقيمين:



تبدو مخطط دالة sigmoid كما يلي:



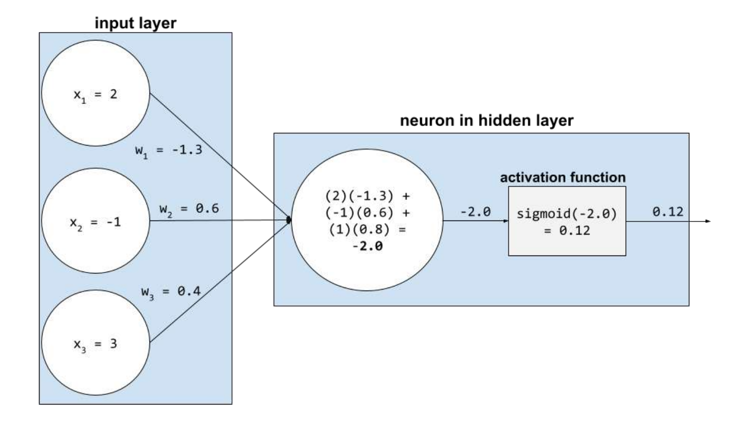
في الشبكة العصبية، تتلاعب دوال التنشيط بالمجموع المرجح (weighted sum) لجميع مدخلات الخلايا العصبية (neuron). لحساب مجموع مرجح، تجمع الخلايا العصبية (neuron) نواتج القيم والأوزان ذات الصلة. على سبيل المثال، افترض أن المدخلات ذات الصلة بالخلايا العصبية تتكون مما يلي:

input value	input weight
2	-1.3
-1	0.6
3	0.4

وبالتالي فإن المجموع المرجح هو:

```
weighted sum = (2)(-1.3) + (-1)(0.6) + (3)(0.4) = -2.0
```

لنفترض أن مصمم هذه الشبكة العصبية اختار دالة sigmoid لتكون دالة التنشيط.في هذه الحالة، تحسب الخلية العصبية sigmoid لـ 2.0-، وهو ما يقرب من 0.12. لذلك، تمر الخلية العصبية 0.12 (بدلاً من 2.0-) إلى الطبقة التاليةفي الشبكة العصبية. يوضح الشكل التالي الجزء ذي الصلة من العملية:



Active learning

التعلم النشط

نهج تدريبي (training) تختار فيه الخوارزمية بعض البيانات التي تتعلم منها. التعلم النشط ذو قيمة خاصة عندما تكون الأمثلة المصنفة نادرة أو مكلفة للحصول عليها. بدلاً من البحث الأعمى عن مجموعة متنوعة من الأمثلة المصنفة (labeled examples)، تبحث خوارزمية التعلم النشط بشكل انتقائي عن مجموعة معينة من الأمثلة التي تحتاجها للتعلم.

AdaDelta

Adadelta هو امتداد أكثر قوة لـ Adagrad يقوم بتكييف معدلات التعلم (learning rates) بناءً على نافذة متحركة لتحديثات التدرج (gradient) ، بدلاً من تجميع جميع التدرجات السابقة. بهذه الطريقة، يستمر Adadelta في التعلم حتى بعد إجراء العديد من التحديثات.

AdaGrad

خوارزمية تدرج عشوائي (gradient descent) متطورة تعيد قياس تدرجات كل معلمة (parameter)، مما يمنح كل معلمة معدل تعلم (learning rate) مستقل بشكل فعال.

Adaptive Learning Rate Algorithms

خوارزمية معدل التعلم التكيفي

تعد طرق معدل التعلم التكيفي بمثابة تحسين (optimization) لطرق التدرج الاشتقاقي (gradient descent) بهدف تقليل دالة الهدف (objective function) للشبكة باستخدام تدرج الدالة ومعلمات الشبكة.

Adaptive Moment Estimation Algorithm (Adam)

آدم

آدم (Adam) هو طريقة أخرى تحسب معدلات التعلم التكيفي لكل معلمة. بالإضافة إلى تخزين متوسط التناقص الأسي للتدرجات التربيعية السابقة مثل Adadelta و RMSprop ، يحتفظ آدم أيضًا بمتوسط متدهور بشكل كبير للتدرجات السابقة v ، على غرار الزخم (momentum).

Adversarial Learning Strategy

استراتيجية التعلم العدائي

التعلم الآلي العدائي هو أسلوب يستخدمفي التعلم الآلي (machine learning) لخداع نموذج أو تضليله بمدخلات ضارة. بينما يمكن استخدام التعلم الآلي العدائيفي مجموعة متنوعة من التطبيقات، فإن هذه التقنية هي الأكثر استخدامًا لتنفيذ هجوم أو التسببفي حدوث خللفي نظام التعلم الآلي.

Agent

الوكيل

في التعلم المعزز (reinforcement learning)، الكيان الذي يستخدم سياسة (policy) لتعظيم العائد (return) المتوقع المكتسب من الانتقال بين حالات (states) البيئة (environment).

Agglomerative clustering

التكتل التراكمي

انظر التكتل الهرمي (hierarchical clustering).

Anchor Boxes

مربعات الارتكاز

مربعات الارتكاز هي مجموعة من الصناديق المحيطة المحددة مسبقًا بارتفاع وعرض معين. يتم تحديد هذه المربعات لالتقاط المقياس ونسبة العرض إلى الارتفاع لفئات الكائنات المحددة التي تريد اكتشافها ويتم اختيارها عادةً بناءً على أحجام العناصرفي مجموعات بيانات التدريب الخاصة بك.

Anomaly detection

اكتشاف الانحراف (القيم المتطرفة)

عملية تحديد القيم المتطرفة (outliers). على سبيل المثال، إذا كان المتوسط لميزة معينة هو 100 بانحراف معياري قدره 10، فيجب أن يشير اكتشاف الانحراف (anomaly detection) إلى قيمة 200 على أنها مشبوهة.

Apache Spark

Apache Spark هو إطار عمل للحوسبة العنقودية مفتوح المصدر. يمكن نشر Spark بعدة طرق، ويوفر روابط أصلية للغات برمجة Java و Scala و Python و R، ويدعم SQL وتدفق البيانات والتعلم الآلي.

AR

اختصار للواقع المعزز (augmented reality).

Area under the PR curve

انظر PR AUC (Area under the PR Curve)

Area under the ROC curve

انظر AUC (Area under the ROC curve)

Artificial intelligence

الذكاء الاصطناعي

برنامج أو نموذج (model) غير بشري يمكنه حل المهام المعقدة. على سبيل المثال، يُظهر البرنامج أو النموذج الذي يترجم نصًا أو برنامجًا أو نموذجًا يحدد الأمراض من الصور الإشعاعية ذكاءً اصطناعيًا.

بشكل رسمي، يعد التعلم الآلي (machine learning) مجالًا فرعيًا للذكاء الاصطناعي. ومع ذلك،في السنوات الأخيرة، بدأت بعض المنظماتفي استخدام مصطلحات الذكاء الاصطناعي والتعلم الآلي بالتبادل.

Attention

الانتباه

أي مجموعة واسعة من آليات بنية الشبكة العصبية (neural network) التي تجمع المعلومات من مجموعة من المدخلات بطريقة تعتمد على البيانات. قد تتكون آلية الانتباه النموذجية من مجموع مرجح على مجموعة من المدخلات، حيث يتم حساب الوزن (weight) لكل مدخل بواسطة جزء آخر من الشبكة العصبية.

راجع أيضًا إلى الانتباه الذاتي (self-attention) والانتباه الذاتي متعدد الرؤوس (multi-head self-attention)، وهما اللبنات الأساسية للمحولات (Transformers).

Attribute

السمة

مرادف للميزة (feature).

في التعلم الآلي، غالبًا ما تشير السمات إلى الخصائص المتعلقة بالأفراد.

Attribute sampling

أخذ عينات السمة

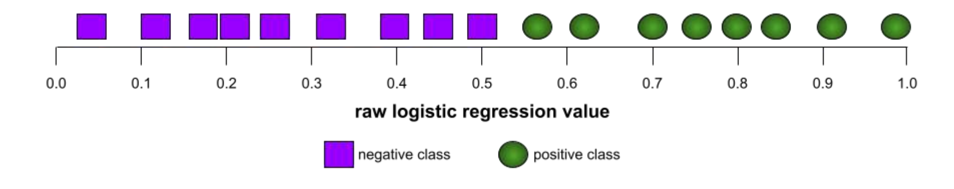
تكتيك لتدريب مجموعة قرارات (decision forest) تأخذ فيها كل شجرة قرار (decision tree) بعين الاعتبار مجموعة فرعية عشوائية من الميزات المحتملة عند تعلم الحالة (condition). بشكل عام، يتم أخذ عينات من مجموعة فرعية مختلفة من الميزات لكل عقدة (node).في المقابل، عند تدريب شجرة قرار بدون أخذ عينات السمة، يتم أخذ جميع الميزات الممكنةفي الاعتبار لكل عقدة.

AUC (Area under the ROC curve)

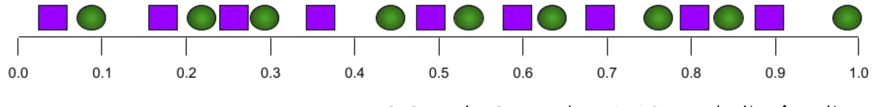
AUC (المنطقة الواقعة تحت منحنى ROC)

رقم بين 0.0 و1.0 يمثل قدرة نموذج التصنيف الثنائي على فصل الفئات الإيجابية عن الفئات السلبية. كلما اقتربت AUC من 1.0، زادت قدرة النموذج على فصل الفئات عن بعضها البعض.

على سبيل المثال، يوضح الرسم التوضيحي التالي نموذج مصنف يفصل بشكل مثالي الفئات الإيجابية (الأشكال البيضاوية الخضراء) عن الفئات السلبية (المستطيلات الأرجواني). هذا النموذج المثالي غير الواقعي لديه AUC يساوي 1.0:

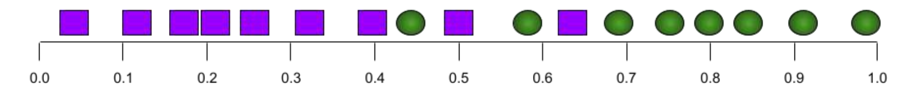


على العكس من ذلك، يوضح الرسم التوضيحي التالي نتائج نموذج المصنف الذي أدى إلى نتائج عشوائية. هذا النموذج لديه AUC من 0.5:



نعم، النموذج السابق به AUC يساوي 0.5 وليس 0.0.

تقع معظم النماذج في مكان ما بين الطرفين. على سبيل المثال، يفصل النموذج التالي الإيجابيات عن السلبيات إلى حد ما، وبالتالي يحتوي على AUC في مكان ما بين 0.5 و1.0:



تتجاهل AUC أي قيمة تحددها لعتبة التصنيف (classification threshold). بدلاً من ذلك، تنظر AUC في جميع عتبات التصنيف الممكنة.

Augmented reality

الواقع المعزز

تقنية تقوم بتركيب صورة تم إنشاؤها بواسطة الكمبيوتر على رؤية المستخدم للعالم الحقيقي، وبالتالي توفير عرض مركب.

AutoEncoder

المشفر التلقائي

يعد المشفر التلقائي نوعًا من الشبكات العصبية الاصطناعية (artificial neural network) المستخدمة لتعلم الترميز الفعال للبيانات غير المصنفة (التعلم غير الخاضع للإشراف unsupervised learning). يتم التحقق من صحة الترميز وصقله من خلال محاولة إعادة إنشاء المدخلات من الترميز. يتعلم المشفر التلقائي تمثيلاً (ترميزًا) لمجموعة من البيانات، عادةً لتقليل الأبعاد (dimensionality reduction)، عن طريق تدريب الشبكة على تجاهل البيانات غير المهمة ("الضوضاء noise").

Automatic Differentiation

التفاضل التلقائي

التفاضل التلقائي (AD) عبارة عن مجموعة من التقنيات لتحويل برنامج يحسب القيم العددية لدالة ما، إلى برنامج يحسب القيم العددية لمشتقات تلك الدالة بنفس الدقة والكفاءة تقريبًا مثل قيم الدالة نفسها.

Auxiliary Loss

الخطأ الإضافي

الخطأ المساعد هو خطأ إضافي إلى جانب خطأ الفرع الرئيسي للمساعدةفي تحسين عملية التعلم للشبكات العصبية (neural networks).

Average Pooling

التجميع المتوسط

التجميع المتوسط هو عملية تجميع تحسب متوسط قيمة تصحيحات خريطة المعالم (feature map)، وتستخدمها لإنشاء خريطة معالم مختزلة (downsampled) (مجمعة pooled). يستخدم عادة بعد الطبقة التلافيفية (convolutional layer).

Average precision

متوسط الدقة

مقياس لتلخيص أداء سلسلة مرتبة من النتائج. يتم حساب متوسط الدقة بأخذ متوسط قيم الدقة لكل نتيجة ذات صلة (كل نتيجةفي القائمة المرتبة حيث يزيد الاستدعاء (recall) بالنسبة إلى النتيجة السابقة).

انظر أيضًا Area under the PR Curve.

Axis-aligned condition

شرط محاذاة المحور

في شجرة القرار (decision tree)، شرط (condition) يتضمن ميزة (feature) واحدًا فقط. على سبيل المثال، إذا كانت المنطقة ميزة، فإن ما يلي هو شرط محاذاة المحور:

```
area > 200
```

B

B

Backpropagation

الانتشار الخلفي

الخوارزمية التي تنفذ التدرج الاشتقاقي (gradient descent)في الشبكات العصبية (neural networks).

يتضمن تدريب الشبكة العصبية العديد من التكرارات (iterations) لدورة المسارين التالية:

1. أثناء التمرير الأمامي (forward pass)، يعالج النظام دفعة (batch) من الأمثلة للحصول على تنبؤ (تنبؤات predictions). يقارن النظام كل توقع بكل قيمة تسمية (label). الفرق بين التنبؤ وقيمة التسمية هو الخطأ (loss) لهذا المثال. يقوم النظام بتجميع الاخطاء لجميع الأمثلة لحساب الخطأ الإجمالي للدُفعة الحالية.
2. أثناء التمرير الخلفي (backward pass) او الانتشار الخلفي (backpropagation)، يقلل النظام من الخطأ عن طريق ضبط أوزان (weights) جميع الخلايا العصبية (neurons)في جميع الطبقات المخفية (hidden layer).

غالبًا ما تحتوي الشبكات العصبية على العديد من الخلايا العصبية عبر العديد من الطبقات المخفية. يساهم كل من هذه الخلايا العصبيةفي الخطأ الكلي بطرق مختلفة. يحدد الانتشار الخلفي (Backpropagation) ما إذا كان يجب زيادة أو تقليل الأوزان المطبقة على خلايا عصبية معينة.

معدل التعلم (learning rate) هو معلمة فائقة (hyperparameter) يتحكمفي الدرجة التي يزيد بها كل تمريرة للخلف أو ينقص كل وزن. معدل التعلم الكبير سيزيد أو ينقص كل وزن أكثر من معدل التعلم الصغير.

من حيث حساب التفاضل والتكامل، يطبق الانتشار الخلفي (backpropagation)قاعدة السلسلة (chain rule) حساب التفاضل والتكامل. أي أن الانتشار الخلفي يحسب المشتق الجزئي للخطأ فيما يتعلق بكل معلمة. لمزيد من التفاصيل.

منذ سنوات، كان على ممارسي التعلم الآلي كتابة كود لتنفيذ الانتشار الخلفي. تقوم واجهات برمجة تطبيقات التعلم الآلي الحديثة مثل TensorFlow الآن بتنفيذ الانتشار الخلفي نيابة عنك.

Bagging

التعبئة

التعبئة، المعروف أيضًا باسم تجميع التمهيد (bootstrap aggregation)، هو طريقة تعلم المجموعة المستخدمة بشكل شائع لتقليل التباين (variance) داخل مجموعة البيانات الصاخبة (noisy data).في التعبئة، يتم اختيار عينة عشوائية من البياناتفي مجموعة التدريب مع الاستبدالــ مما يعني أنه يمكن اختيار نقاط البيانات الفردية أكثر من مرة.

Bag of words (BOW)

حقيبة الكلمات

تمثيل الكلماتفي عبارة أو مقطع، بغض النظر عن الترتيب. على سبيل المثال، تمثل حقيبة الكلمات العبارات الثلاث التالية بشكل متماثل:

- the dog jumps
- jumps the dog
- dog jumps the

يتم تعيين كل كلمة إلى فهرسفي متجه متناثر (sparse vector)، حيث يحتوي المتجه على فهرس لكل كلمةفي المفردات. على سبيل المثال، يتم تعيين العبارة "the dog jumps"في متجه ميزة (feature vector) بقيم غير صفريةفي المؤشرات الثلاثة المقابلة للكلمات the، dog، jumps. يمكن أن تكون القيمة غير الصفرية أيًا مما يلي:

a) 1 للدلالة على وجود كلمة.

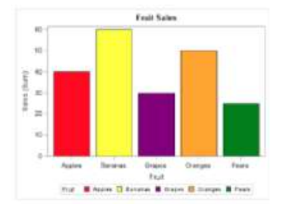
b) عدد مرات ظهور كلمةفي الحقيبة. على سبيل المثال، إذا كانت العبارة " the maroon dog" هي "dog with maroon fur ”، فسيتم تمثيل كل من maroon و dogبالرقم 2، بينما سيتم تمثيل الكلمات الأخرى بالرمز 1.

c) قيمة أخرى، مثل لوغاريتم عدد المرات التي تظهر فيها الكلمةفي الحقيبة.

Bar Chart

مخطط شريطي

المخططات الشريطية هي نوع من الرسوم البيانية التي تُستخدم لعرض ومقارنة الأرقام أو التكرار أو المقاييس الأخرى (مثل المتوسط mean) لفئات منفصلة مختلفة من البيانات. يتم استخدامها للمتغيرات الفئوية (categorical variables).



Baseline

خط الاساس

نموذج (model) يُستخدم كنقطة مرجعية لمقارنة مدى جودة أداء نموذج آخر (نموذج أكثر تعقيداً). على سبيل المثال، قد يعمل نموذج الانحدار اللوجستي (logistic regression model) كخط أساس جيد لنموذج عميق (deep model).

بالنسبة لمشكلة معينة، يساعد خط الأساس مطوري النماذج على تحديد الحد الأدنى من الأداء المتوقع الذي يجب أن يحققه النموذج الجديد حتى يكون النموذج الجديد مفيداً.

Batch

الدفعة

مجموعة الأمثلة (examples) المستخدمةفي تكرار (iteration) واحد تدريب. يحدد حجم الدُفعة (batch size) عدد الأمثلةفي الدُفعة.

راجع الحقبة (الفترة) (epoch) للحصول على شرح لكيفية ارتباط الدفعة بحقبة ما

Batch normalization

التسوية بالدفعات

التسوية (Normalizing) إدخال أو إخراج دوال التنشيط (activation functions)في طبقة مخفية (hidden layer). يمكن أن يوفر التسوية بالدفعات الفوائد التالية:

- جعل الشبكات العصبية (neural networks) أكثر استقرارًا من خلال الحماية من الأوزان المتطرفة (outlier).
- تمكين معدلات التعلم (learning rates) العالي، والتي يمكن أن تسرع التدريب.
- تقليل الضبط الزائد (overfitting).

Batch size

الدفعة

عدد الأمثلة (examples)في الدفعة(batch). على سبيل المثال، إذا كان حجم الدُفعة هو 100، فإن النموذج يعالج 100 مثال لكل تكرار (iteration).

فيما يلي استراتيجيات حجم الدُفعة الشائعة:

- التدرج الاشتقاقي العشوائي (Stochastic Gradient Descent (SGD)، حيث يكون حجم الدُفعة 1.
- الدُفعة الكاملة، حيث يكون حجم الدُفعة هو عدد الأمثلةفي مجموعة التدريب (training set) بأكملها. على سبيل المثال، إذا كانت مجموعة التدريب تحتوي على مليون مثال، فسيكون حجم الدفعة مليون مثال. الدفعة الكاملة عادة ما تكون استراتيجية غير فعالة.
- الدُفعة الصغيرة (mini-batch) التي يتراوح حجم الدُفعة فيها عادةً بين 10 و1000. الدُفعة الصغيرة هي عادةً الإستراتيجية الأكثر فعالية.

Bayes Theorem

نظرية بايز

تستخدم نظرية بايز لحساب الاحتمال الشرطي (Conditional probability). الاحتمال الشرطي هو احتمال وقوع حدث "ب" نظرًا لأن الحدث ذي الصلة "أ" قد حدث بالفعل.

Bayesian neural network

شبكة بايزي العصبية

شبكة عصبية احتمالية (neural network) مسؤولة عن عدم اليقينفي الأوزان (weights) والمخرجات. يتنبأ (predicts) نموذج الانحدار القياسي للشبكة العصبية عادةً بقيمة عددية؛ على سبيل المثال، يتوقع أحد النماذج سعر منزل يبلغ 853000.في المقابل، تتنبأ شبكة بايزي العصبية بتوزيع القيم؛ على سبيل المثال، يتوقع أحد النماذج سعر منزل يبلغ 853000 مع انحراف معياري قدره 67200. تعتمد شبكة بايزي العصبية على نظرية بايز (Bayes' Theorem) لحساب أوجه عدم اليقينفي الأوزان والتنبؤات. يمكن أن تكون شبكة بايزي العصبية مفيدة عندما يكون من المهم تحديد مقدار عدم اليقين، كما هو الحالفي النماذج المتعلقة بالمستحضرات الصيدلانية. يمكن لشبكات بايزي العصبية أن تساعد أيضًافي منع الضبط الزائد (overfitting).

Bayesian optimization

تحسين بايزي

تقنية نموذج الانحدار الاحتمالي (probabilistic regression model) لتحسين دوال الهدف (objective functions) مكلفة حسابيًا عن طريق تحسين البديل الذي يحدد مقدار عدم اليقين عبر تقنية التعلم بايزي. نظرًا لأن تحسين بايزي مكلف للغايةفي حد ذاته، فإنه يتم استخدامه عادةً لتحسين المهام المكلفة للتقييم التي تحتوي على عدد صغير من المعلمات، مثل اختيار المعلمات الفائقة (hyperparameters).

Bellman equation

معادلة بيلمان

في التعلم المعزز (reinforcement learning)، تكتفي الهوية التالية بدالة Q (Q-function) المثلى:

$$Q(s,a) = r(s,a) + \gamma \mathbb{E}_{s'|s,a} \max_{a'} Q(s',a')$$

تطبق خوارزميات التعلم المعزز هذه الهوية لإنشاء (Q-learning) عبر قاعدة التحديث التالية:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r(s,a) + \gamma \max_{a_1} Q(s',a') - Q(s,a) \right]$$

بالإضافة إلى التعلم المعزز، فإن لمعادلة بيلمان تطبيقاتفي البرمجة الديناميكية.

BERT (Bidirectional Encoder Representations from Transformers)

بيرت

معمارية نموذج لتمثيل (representation) النص. يمكن أن يعمل نموذج BERT المدرب كجزء من نموذج أكبر لتصنيف النص أو مهام التعلم الآلي الأخرى.

يتميز BERT بالخصائص التالية:

- يستخدم معمارية المحول (Transformer)، وبالتالي يعتمد على الانتباه الذاتي (self-attention).
- يستخدم جزء المشفر (encoder) من المحول. وظيفة المشفر هي إنتاج تمثيلات نصية جيدة، بدلاً من أداء مهمة محددة مثل التصنيف.
- ثنائي الاتجاه (bidirectional).
- يستخدم إخفاء (masking) للتدريب غير الخاضع للأشراف (unsupervised training).

تشمل متغيرات BERT ما يلي:

- ALBERT ، وهو اختصار لـ A Light BERT.
- LaBSE.

Bias

التحيز

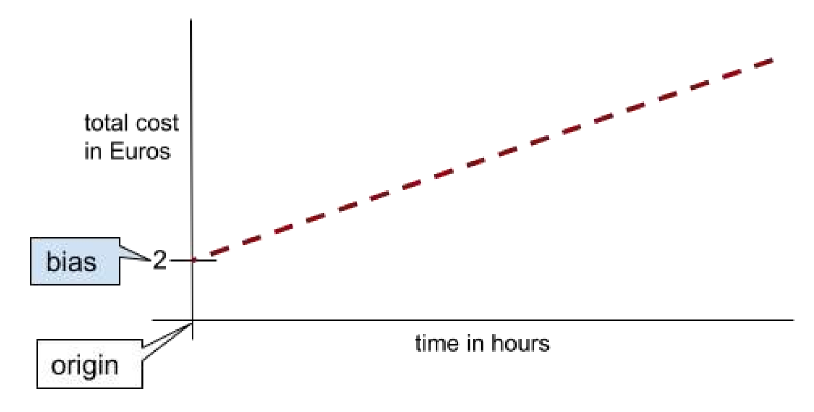
اعتراض أو إزاحة من نقطة الأصل. التحيز هو معلمةفي نماذج التعلم الآلي، والتي يرمز إليها بأي مما يلي:

- *b*
- w_0

على سبيل المثال، التحيز هو *b*في الصيغة التالية:

$$y' = b + w_1 x_1 + w_2 x_2 + \ldots w_n x_n$$

في خط ثنائي الأبعاد بسيط، يعني التحيز "تقاطع y". على سبيل المثال، تحيز السطرفي الرسم التوضيحي التالي هو 2.



يوجد التحيز لأنه لا تبدأ جميع النماذج من نقطة الأصل (0،0). على سبيل المثال، افترض أن متنزهًا يكلف 2 يورو للدخول و0.5 يورو إضافي لكل ساعة يقضيها العميل. لذلك، فإن نموذج تعيين التكلفة الإجمالية له تحيز 2 لأن أقل تكلفة هي 2 يورو.

Bias metric

مقياس التحيز

ما هو متوسط الفرق بين توقعاتك والقيمة الصحيحة لتلك المشاهدة (observation)؟

- قد يعني التحيز المنخفض (Low bias) أن كل تنبؤ صحيح. قد يعني أيضًا أن نصف توقعاتك أعلى من قيمها الفعلية ونصفها أقل، بنسب متساوية، مما يؤدي إلى انخفاض متوسط الفرق.
- يشير التحيز المرتفع (High bias) (مع تباين منخفض (low variance)) إلى أن نموذجك قد يكون غير ملائم وأنك تستخدم بنية خاطئة للوظيفة.

Bias-Variance Trade-off

مقايضة التباين التحيز

يعني خطأ التحيز المرتفع (high bias) أن لدينا نموذجًا ضعيف الأداء يحافظ على فقدان الاتجاهات المهمة. سوف يتناسب نموذج التباين العالي (high variance) مع مجتمع التدريب لديك ويؤدي بشكل سيئ في أي ملاحظة بعد التدريب. من أجل الحصول على توافق مثالي في النموذج، يجب أن يكون التحيز والتباين متوازنين وهو ما يمثل مقايضة التباين_التحيز.

Big Data

البيانات الضخمة

البيانات الضخمة هي مصطلح يصف الحجم الكبير من البيانات — المهيكلة (structured) وغير المهيكلة (unstructured). لكن ليس حجم البيانات هو المهم. إنها الطريقة التي تستخدم بها المؤسسات هذا الكم الهائل من البيانات لإنشاء الإحصاءات. تستخدم الشركات أدوات وتقنيات وموارد مختلفة لفهم هذه البيانات لاشتقاق استراتيجيات عمل فعالة.

Bigram

Bigram هو N-gram فيه N=2.

Bidirectional

ثنائي الاتجاه

مصطلح يستخدم لوصف نظام يقوم بتقييم النص الذي يسبق ويتبع القسم الهدف من النص.في المقابل، يقوم النظام أحادي (unidirectional) الاتجاه فقط بتقييم النص الذي يسبق القسم الهدف من النص.

على سبيل المثال، ضع في اعتبارك نموذج لغة مقنعة (masked language model) يجب أن يحدد احتمالات الكلمة أو الكلمات التي تمثل التسطير في السؤال التالي:

What is the _____ with you?

يجب أن يبني نموذج اللغة أحادي الاتجاه احتمالاته فقط على السياق الذي توفره الكلمات "What" و "is" و "the". في المقابل، يمكن أن يكتسب نموذج اللغة ثنائي الاتجاه أيضًا سياق من "with" و "you"، مما قد يساعد النموذج في إنشاء تنبؤات أفضل.

Bidirectional language model

نموذج لغة ثنائي الاتجاه

نموذج لغوي (language model) يحدد احتمال وجود رمز معين في مكان معين في مقتطف من النص بناءً على النص السابق والتالي.

Binary classification

التصنيف الثنائي

نوع من مهام التصنيف (classification) التي تتنبأ بواحد من فئتين متعارضتين:

- الفئة الإيجابية (positive class).
- الفئة السلبية (negative class).

على سبيل المثال، يقوم كل من نموذجي التعلم الآلي التاليين بتصنيف ثنائي:

- نموذج يحدد ما إذا كانت رسائل البريد الإلكتروني هي بريد عشوائي (الفئة الإيجابية) أم ليست بريداً عشوائيًا (الفئة السلبية).
- نموذج يقيم الأعراض الطبية لتحديد ما إذا كان الشخص مصابًا بمرض معين (الفئة الإيجابية) أو لا يعاني من هذا المرض (الفئة السلبية).

على النقيض من التصنيف متعدد الفئات (multi-class classification).

انظر أيضًا الانحدار اللوجستي (logistic regression) وعتبة التصنيف (classification threshold).

Binary Cross-Entropy

الانتروبيا المتقاطعة الثنائية

تقارن الانتروبيا المتقاطعة الثنائية بين كل من الاحتمالات المتوقعة بمخرجات الفئة الفعلية والتي يمكن أن تكون إما 0 أو 1. ثم تحسب الدرجة التي تعاقب الاحتمالات بناءً على المسافة من القيمة المتوقعة. هذا يعني مدى قرب أو بعد القيمة الفعلية.

Binary condition

الشرط الثنائي

في شجرة القرار (decision tree)، شرط (condition) له نتيجتان محتملتان فقط، عادةً نعم أو لا. على سبيل المثال، ما يلي هو شرط ثنائي:

```
temperature >= 100
```

Binning

مرادف لـ bucketing.

BLEU (Bilingual Evaluation Understudy)

دراسة تقييم ثنائية اللغة

درجة بين 0.0 و1.0، شاملة، تشير إلى جودة الترجمة بين لغتين بشريتين (على سبيل المثال، بين الإنجليزية والروسية). تشير درجة BLEU البالغة 1.0 إلى ترجمة مثالية؛ درجة BLEU 0.0 تشير إلى ترجمة سيئة.

Boosting

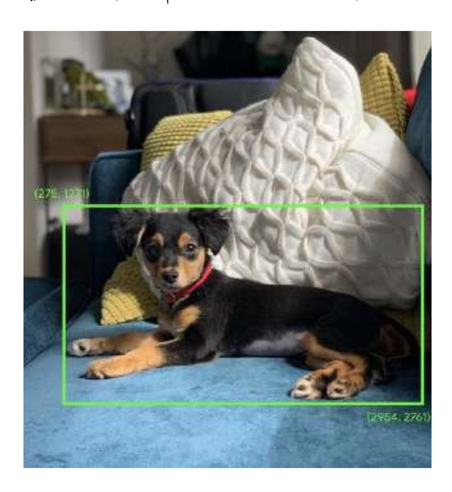
التعزيز

أسلوب التعلم الآلي الذي يجمع بشكل متكرر مجموعة من المصنفات البسيطة وغير الدقيقة للغاية (يشار إليها باسم المصنفات "الضعيفة")في مصنف بدقة عالية (مصنف "قوي") عن طريق زيادة وزن (upweighting) الأمثلة التي يخطئ النموذج حاليًافي تصنيفها.

Bounding box

مربع الإحاطة

في صورة ما، إحداثيات (x, y) لمستطيل حول منطقة اهتمام، مثل الكلبفي الصورة أدناه.



Broadcasting

البث

توسيع شكل المعاملفي عملية الرياضيات المصفوفة إلى أبعاد (dimensions) متوافقة لتلك العملية. على سبيل المثال، يتطلب الجبر الخطي أن المعاملينفي عملية إضافة مصفوفة يجب أن يكون لهما نفس الأبعاد. وبالتالي، لا يمكنك إضافة مصفوفة من الشكل (m, n) إلى متجه طوله

n. يتيح البث هذه العملية عن طريق توسيع متجه الطول n فعليًا إلى مصفوفة من الشكل (m, n) ، عن طريق تكرار نفس القيم أسفل كل عمود.

على سبيل المثال، بالنظر إلى التعريفات التالية، يحظر الجبر الخطي A + B لأن A و B لهما أبعاد مختلفة:

```
A = [[7, 10, 4],
    [13, 5, 9]]
B = [2]
```

ومع ذلك، فإن البث يمكّن العملية A + B من خلال توسيع B افتراضيًا إلى:

```
[[2, 2, 2],
 [2, 2, 2]]
```

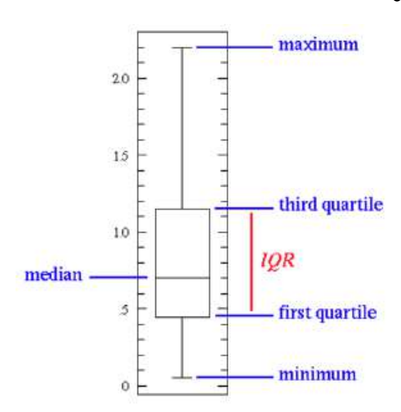
وبالتالي، أصبحت A + B الآن عملية صالحة:

```
[[7, 10, 4],  +  [[2, 2, 2],  =  [[ 9, 12, 6],
 [13, 5, 9]]      [2, 2, 2]]      [15, 7, 11]]
```

Box Plot

المخطط الصندوقي

يعرض النطاق الكامل للتباين (من الحد الأدنى إلى الحد الأقصى)، والنطاق المحتمل للتباين (النطاق بين الشرائح الربعية (Interquartile range))، والقيمة النموذجية (الوسيط median). يوجد أدناه تصور لمخطط الصندوق:



بعض الاستدلالات التي يمكن استنتاجها من مخطط الصندوق:

- الوسيط (Median): يمثل الربع الأوسط الوسيط.
- المربع الأوسط (Middle box): يمثل 50٪ من البيانات.
- الربع الأول (First quartile): 25٪ من البيانات تقع تحت هذا الخط

- الربع الثالث (Third quartile): 75٪ من البيانات تقع تحت هذا الخط.

Bucketing

الحاوية

تحويل ميزة (feature) واحدة إلى ميزات ثنائية متعددة تسمى (buckets) أو (bins)، وعادةً ما تستند إلى نطاق القيمة. عادةً ما تكون ميزة التقطيع ميزة مستمرة (continuous feature).

على سبيل المثال، بدلاً من تمثيل درجة الحرارة كميزة واحدة مستمرة للفاصلة العائمة، يمكنك تقطيع نطاقات درجات الحرارة إلى حاويات (دلو) مقطعة، مثل:

- <= 10 درجات مئوية ستكون الحاوية "البارد".
- 11 ـ 24 درجة مئوية ستكون الحاوية "المعتدل".
- = 25 درجة مئوية ستكون الحاوية "الدافئ".

سيعامل النموذج كل قيمةفي نفس المجموعة بشكل متماثل. على سبيل المثال، القيمتان 13 و22 كلاهمافي حاوية درجة الحرارة المعتدلة، لذا يتعامل النموذج مع القيمتين بشكل متماثل.

C

C

Caffe

Caffe هو إطار عمل تعليمي عميق، تم تطويره في الأصل في جامعة كاليفورنيا ، بيركلي. إنه مفتوح المصدر، بموجب ترخيص BSD. تمت كتابته بلغة C ++، مع واجهة Python .

Calibration layer

طبقة المعايرة

تعديل لاحق للتنبؤ، عادة لحساب تحيز التنبؤ (prediction bias). يجب أن تتطابق التوقعات والاحتمالات المعدلة مع توزيع مجموعة التسميات المرصودة.

Candidate generation

توليد المرشح

المجموعة الأولية من التوصيات المختارة من قبل نظام التوصية (recommendation system). على سبيل المثال، ضع في اعتبارك متجر كتب يقدم 100000 عنوان. تُنشئ مرحلة توليد المرشح قائمة أصغر بكثير من الكتب المناسبة لمستخدم معين، على سبيل المثال 500 كتاب. ولكن حتى 500 كتاب تعد كثيرة جدًا بحيث لا يمكن التوصية بها للمستخدم. المراحل اللاحقة والأكثر تكلفة لنظام التوصيات (مثل تسجيل النقاط (scoring) وإعادة الترتيب (re-ranking)) تقلل من تلك 500 إلى مجموعة توصيات أصغر وأكثر فائدة.

Candidate sampling

أخذ العينات المرشح

تحسين وقت التدريب حيث يتم حساب الاحتمالية لجميع التسميات الإيجابية (positive labels)، باستخدام، على سبيل المثال، softmax، ولكن فقط لعينة عشوائية من التسميات السلبية (negative labels). على سبيل المثال، إذا كان لدينا مثال معنون بأخذ عينات من البيجل والكلاب يحسب الاحتمالات المتوقعة وشروط الخطأ المقابلة لمخرجات فئة البيجل والكلب بالإضافة إلى مجموعة فرعية عشوائية من الفئات المتبقية (القط، المصاصة، السياج). الفكرة هي أن الطبقات السلبية يمكن أن تتعلم من التعزيز السلبي الأقل تكرارًا طالما أن الطبقات الإيجابية (positive classes) تحصل دائمًا على تعزيز إيجابي مناسب، وهذا بالفعل لوحظ تجريبيًا. الدافع وراء أخذ العينات المرشح هو الفوز في الكفاءة الحسابية من عدم حساب التنبؤات لجميع السلبيات.

Categorical Cross-Entropy

الانتروبيا المتقاطعة الفئوية

تستخدم كدالة خطأ (loss function) لنموذج التصنيف متعدد الفئات (multi-class classification) حيث يوجد اثنان أو أكثر من تسميات الإخراج.

Categorical data

البيانات الفئوية

الميزات (Features) التي تحتوي على مجموعة محددة من القيم الممكنة. على سبيل المثال، ضعفي اعتبارك ميزة فئوية تسمى حالة إشارة المرور، والتي يمكن أن تحتوي فقط على واحدة من القيم الثلاث المحتملة التالية:

- أحمر.
- الأصفر.
- أخضر.

من خلال تمثيل حالة إشارة المرور كميزة فئوية، يمكن للنموذج معرفة التأثيرات المختلفة للأحمر والأخضر والأصفر على سلوك السائق.

تسمى الميزات الفئوية أحيانًا ميزات متقطعة (discrete features).

على النقيض من البيانات العددية (numerical data).

Causal language model

نموذج اللغة السببية

مرادف لـ (unidirectional language model).

انظر نموذج اللغة ثنائي الاتجاه (bidirectional language model) لمقارنة الأساليب الاتجاهية المختلفةفي نمذجة اللغة.

CBOW

حقيبة الكلمات المستمرّة

في نموذج CBOW، يتم دمج تمثيلات السياق الموزعة (أو الكلمات المحيطة) للتنبؤ بالكلمةفي المنتصف.

Centroid

سنترويد (نقطة المركز)

مركز التكتل (cluster) كما هو محدد بواسطة خوارزمية k-mean أو k-median. على سبيل المثال، إذا كان k هو 3، فإن خوارزمية (k-mean) أو (k-median) تعثر على 3 سنترويد.

Centroid-based clustering

التكتل القائم على السنترويد

فئة من خوارزميات التكتل (clustering) التي تنظم البياناتفي مجموعات غير هرمية. k-mean هي خوارزمية التكتل الأكثر استخدامًا على أساس السنترويد.

على النقيض من خوارزميات المجموعات الهرمية (hierarchical clustering).

Checkpoint

نقطة حفظ

البيانات التي تلتقط حالة معلمات (parameters) النموذجفي تكرار تدريب معين. تتيح نقاط الحفظ تصدير أوزان (weights) النموذج، أو إجراء تدريب (training) عبر جلسات متعددة. تتيح نقاط الحفظ أيضًا التدريب على متابعة أخطاء الماضي (على سبيل المثال، العمل الوقائي).

Class

الفئة

فئة يمكن أن تنتمي إليها التسمية (label). فمثلا:

- في نموذج التصنيف الثنائي (binary classification) الذي يكتشف البريد العشوائي، قد تكون الفئتان عبارة عن بريد عشوائي وليست بريدًا عشوائيًا.
- في نموذج التصنيف متعدد الفئات (multi-class classification) الذي يحدد سلالات الكلاب، قد تكون الفئات هي البودل، والبيغل، والبج، وما إلى ذلك.

يتنبأ نموذج التصنيف (classification model) بفئة.في المقابل، يتنبأ نموذج الانحدار (regression model) برقم بدلاً من فئة.

Classifier

المصنف

تستخدم المصنفات (نماذج التصنيف classification models) للتنبؤ (prediction) بالنتائج الفئوية (categorical).

Classification

التصنيف

التصنيف هو عملية تصنيف مجموعة معينة من البيانات إلى فئات (classes)، ويمكن إجراؤها على كل من البيانات المهيكلة (structured) أو غير المهيكلة (unstructured). تبدأ العملية بالتنبؤ بفئة نقاط البيانات المحددة. غالبًا ما يشار إلى الفئات على أنها الهدف (target) أو التسمية (label) أو الاصناف (categories).

Classification model

نموذج التصنيف

نموذج (model) التنبؤ به فئة. على سبيل المثال، فيما يلي جميع نماذج التصنيف:

- نموذج يتنبأ بلغة جملة الإدخال (الفرنسية؟ الإسبانية؟ الإيطالية؟).
- نموذج يتنبأ بأنواع الأشجار (القيقب، البلوط، الباوباب؟).
- نموذج يتنبأ بالفئة الإيجابية أو السلبية لحالة طبية معينة.

في المقابل، تتنبأ نماذج الانحدار (regression models) بالأرقام بدلاً من الفئات.

هناك نوعان شائعان من نماذج التصنيف هما:

- تصنيف ثنائي (binary classification).
- تصنيف متعدد الفئات (multi-class classification).

Classification threshold

عتبة التصنيف

في التصنيف الثنائي (binary classification)، رقم بين 0 و1 يحول الناتج الخام لنموذج الانحدار اللوجستي (logistic regression) إلى تنبؤ بالفئة الإيجابية (positive class) أو الفئة السلبية (negative class). لاحظ أن عتبة التصنيف هي قيمة يختارها الإنسان، وليست قيمة يختارها تدريب النموذج.

يُخرج نموذج الانحدار اللوجستي قيمة أولية بين 0 و1. ثم:

- إذا كانت هذه القيمة الأولية أكبر من عتبة التصنيف، فسيتم توقع الفئة الإيجابية.

- إذا كانت هذه القيمة الأولية أقل من عتبة التصنيف، فسيتم توقع الفئة السالبة.

على سبيل المثال، افترض أن حد التصنيف هو 0.8. إذا كانت القيمة الأولية 0.9، فإن النموذج يتنبأ بالفئة الإيجابية. إذا كانت القيمة الأولية 0.7، فإن النموذج يتنبأ بالفئة السلبية.

يؤثر اختيار عتبة التصنيف بشدة على عدد الإيجابيات الكاذبة (false positives) والسلبيات الكاذبة (false negatives).

Class-imbalanced dataset

مجموعة بيانات غير متوازنة في الفئة

مجموعة بيانات لمشكلة تصنيف يختلف فيها العدد الإجمالي لتسميات (labels) كل فئة اختلافًا كبيرًا. على سبيل المثال، ضع في اعتبارك مجموعة بيانات التصنيف الثنائية التي يتم تقسيم تسمياتها على النحو التالي:

- 1000000 تسميات سلبية.
- 10 تسميات إيجابية.

نسبة التسميات السلبية إلى الإيجابية هي 100000 إلى 1، لذا فهذه مجموعة بيانات غير متوازنة في الفئة.

في المقابل، فإن مجموعة البيانات التالية ليست غير متوازنة في الفئة لأن نسبة التسميات السلبية إلى التسميات الإيجابية قريبة نسبيًا من 1:

- 517 تسميات سلبية.
- 483 تسميات موجبة.

يمكن أيضًا أن تكون مجموعات البيانات متعددة الفئات غير متوازنة في الفئة. على سبيل المثال، تعتبر مجموعة بيانات التصنيف متعددة الفئات التالية أيضًا غير متوازنة في الفئة لأن أحد التسميات يحتوي على أمثلة أكثر بكثير من الأخريين:

- 1000000 تسمية بفئة "خضراء".
- 200 تسمية بفئة "أرجواني".
- 350 تسمية بفئة "برتقالي".

انظر أيضًا الإنتروبيا (entropy)، وطبقة الأغلبية (majority class)، وطبقة الأقلية (minority class).

Clean Data

البيانات النظيفة

تنظيف البيانات (Data cleaning) هو عملية إصلاح أو إزالة البيانات غير الصحيحة أو التالفة أو المنسقة بشكل غير صحيح أو المكررة أو غير المكتملة ضمن مجموعة البيانات. عند الجمع بين مصادر بيانات متعددة، هناك العديد من الفرص لتكرار البيانات أو تسميتها بشكل خاطئ.

Clipping

القص

تقنية للتعامل مع القيم المتطرفة (outliers) عن طريق القيام بأي من الإجراءين التاليين أو كليهما:

- تقليل قيم الميزة (feature) التي تزيد عن الحد الأقصى إلى الحد الأقصى.
- زيادة قيم الميزة التي تقل عن الحد الأدنى حتى الحد الأدنى.

على سبيل المثال، افترض أن <0.5٪ من القيم لميزة معينة تقع خارج النطاق 40–60.في هذه الحالة، يمكنك القيام بما يلي:

- قص جميع القيم التي تزيد عن 60 (الحد الأقصى) لتكون 60 بالضبط.
- قص جميع القيم التي تقل عن 40 (الحد الأدنى) لتكون 40 بالضبط.

يمكن أن تتسبب القيم المتطرفةفي إتلاف النماذج، مما يؤدي أحيانًا إلى تجاوز الأوزان (weights) أثناء التدريب. يمكن أن تفسد بعض القيم المتطرفة بشكل كبير مقاييس مثل الدقة (accuracy). القص هو أسلوب شائع للحد من الضرر.

يفرض قص التدرج (Gradient clipping) قيم التدرج (gradient) ضمن نطاق معين أثناء التدريب.

Cloud TPU

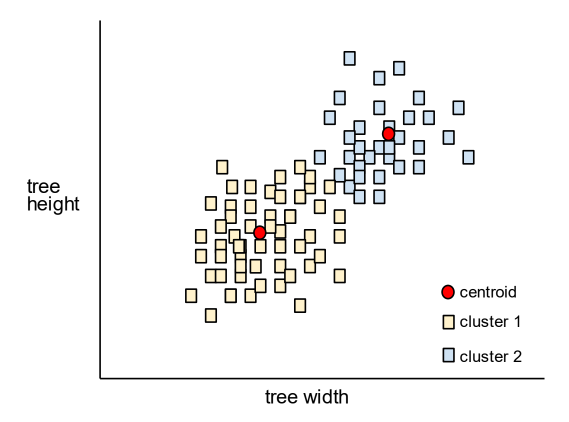
مُسرع أجهزة متخصص مصمم لتسريع أعباء عمل التعلم الآلي على Google Cloud Platform.

Clustering

التكتل

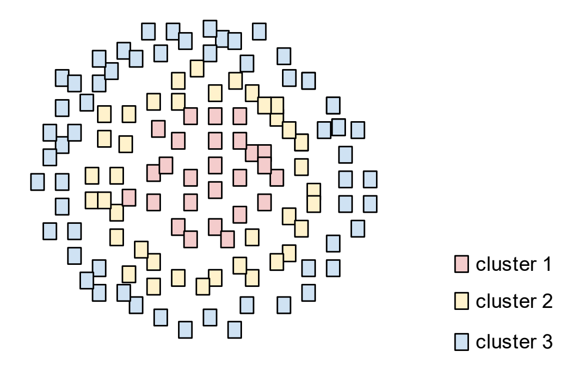
تجميع الأمثلة ذات الصلة، خاصة أثناء التعلم غير الخاضع للإشراف (unsupervised learning). بمجرد أن يتم تجميع جميع الأمثلة، يمكن للإنسان أن يقدم معنى اختياريًا لكل مجموعة.

توجد العديد من خوارزميات التجميع. على سبيل المثال، أمثلة مجموعات خوارزمية (k-mean)بناءً على قربها من النقطة الوسطى (centroid)، كمافي الرسم التخطيطي التالي:



يمكن للباحث البشري بعد ذلك مراجعة المجموعات (clusters)، وعلى سبيل المثال، تسمية المجموعة 1 على أنها "أشجار قزمة" والمجموعة 2 على أنها "أشجار كاملة الحجم".

كمثال آخر، ضعفي اعتبارك خوارزمية التكتل بناءً على مسافة المثال من نقطة المركز، موضحة على النحو التالي:



Co-adaptation

التكيف المشترك

عندما تتنبأ الخلايا العصبية (neurons) بأنماطفي بيانات التدريب من خلال الاعتماد بشكل شبه حصري على مخرجات الخلايا العصبية أخرى محددة بدلاً من الاعتماد على سلوك الشبكة ككل. عندما لا تكون الأنماط التي تسبب التكيف المشترك موجودةفي بيانات التحقق من الصحة، فإن التكيف المشترك يتسببفي الضبط الزائد (overfitting). يقلل تنظيم التسرب (Dropout

(regularization) من التكيف المشترك لأن التسرب يضمن أن الخلايا العصبية لا يمكنها الاعتماد فقط على خلايا عصبية أخرى محددة.

Colab

كولاب

Colaboratory، أو باختصار "Colab"، منتج من Google Research يسمح Colab لأي شخص بكتابة وتنفيذ كود Python من خلال المتصفح ، وهو مناسب بشكل خاص للتعلم الآلي والعميق وتحليل البيانات. بشكل أكثر تقنيًا، Colab عبارة عن خدمة كمبيوتر محمول مستضافة من Jupyter لا تتطلب أي إعداد للاستخدام، مع توفير الوصول المجاني إلى موارد الحوسبة بما في ذلك وحدات معالجة الرسومات GPU.

Collaborative filtering

التصفية التعاونية

عمل تنبؤات (predictions) حول اهتمامات مستخدم واحد بناءً على اهتمامات العديد من المستخدمين الآخرين. غالبًا ما يتم استخدام التصفية التعاونيةفي أنظمة التوصية (recommendation systems).

Computer Vision

الرؤية الحاسوبية

الرؤية الحاسوبية هي أحد مجالات الذكاء الاصطناعي (artificial intelligence) التي تدرب أجهزة الكمبيوتر على تفسير وفهم العالم المرئي. باستخدام الصور الرقمية من الكاميرات ومقاطع الفيديو ونماذج التعلم العميق (deep learning models)، يمكن للآلات تحديد الأشياء وتصنيفها بدقة ـ ثم التفاعل مع ما "تراه".

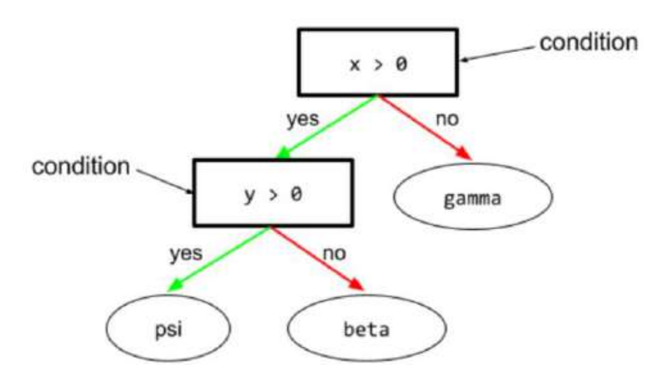
بعض التطبيقات الرئيسية للرؤية الحاسوبية هي:

- اكتشاف المشاة والسيارات والطريقفي السيارات الذكية (ذاتية القيادة self-driving).
- التعرف على الأشياء (Object recognition).
- تتبع الكائن (Object tracking).
- تحليل الحركة (Motion analysis).
- استعادة الصورة (Image restoration).

Condition

الشرط

في شجرة القرار (decision tree)، أي عقدة (node) تقوم بتقييم تعبير. على سبيل المثال، يحتوي الجزء التالي من شجرة القرار على شرطين:



الشرط تسمى أيضًا انقسام أو اختبار.

يتباين الشرط مع الورقة (leaf).

Confusion matrix

مصفوفة الارتباك

جدول NxN يلخص عدد التنبؤات الصحيحة وغير الصحيحة التي قدمها نموذج التصنيف (classification model). على سبيل المثال، ضعفي اعتبارك مصفوفة الارتباك التالية لنموذج التصنيف الثنائي (binary classification):

	Tumor (predicted)	Non-Tumor (predicted)
Tumor (ground truth)	18 (TP)	1 (FN)
Non-Tumor (ground truth)	6 (FP)	452 (TN)

توضح مصفوفة الارتباك السابقة ما يلي:

- من بين التنبؤات الـ 19 التي كانت الحقيقة الأساسية (ground truth) فيها هي الورم (Tumor)، صنف النموذج 18 بشكل صحيح وصنف 1 بشكل غير صحيح.
- من بين 458 تنبوًا كانت الحقيقة الأساسية فيها غير ورم (Non-Tumor)، صنف النموذج بشكل صحيح 452 وصنف 6 بشكل غير صحيح.

يمكن أن تساعدك مصفوفة الارتباك الخاصة بمشكلة التصنيف متعدد الفئات (multi-class classification)في تحديد أنماط الأخطاء. على سبيل المثال، ضعفي اعتبارك مصفوفة الارتباك التالية لنموذج تصنيف متعدد الفئات من 3 فئات يصنف ثلاثة أنواع مختلفة من زهرة iris (فيرجينيكا وفيرسيكولور وسيتوسا). عندما كانت الحقيقة الأساسية هي فيرجينيكا، تُظهر مصفوفة الارتباك أن النموذج كان أكثر عرضة للتنبؤ بفيرسيكولور بالخطأ من سيتوسا:

	Setosa (predicted)	Versicolor (predicted)	Virginica (predicted)
Setosa (ground truth)	88	12	0
Versicolor (ground truth)	6	141	7
Virginica (ground truth)	2	27	109

كمثال آخر، يمكن أن تكشف مصفوفة الارتباك أن نموذجًا تم تدريبه على التعرف على الأرقام المكتوبة بخط اليد يميل إلى التنبؤ بالخطأ 9 بدلاً من 4، أو التنبؤ بالخطأ 1 بدلاً من 7.

تحتوي مصفوفات الارتباك على معلومات كافية لحساب مجموعة متنوعة من مقاييس الأداء، بمافي ذلك الدقة (precision) والاسترجاع (recall).

Continuous feature

الميزة المستمرة

ميزة فاصلة عائمة بمدى لانهائي من القيم الممكنة، مثل درجة الحرارة أو الوزن.

على النقيض من ميزة متقطعة (discrete feature).

Convenience sampling

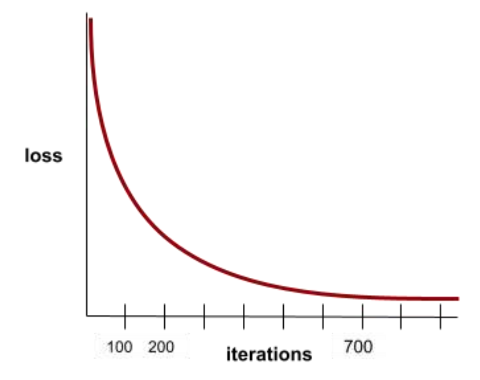
أخذ العينات الملائمة

استخدام مجموعة بيانات لم يتم جمعها علميًا لإجراء تجارب سريعة.في وقت لاحق، من الضروري التبديل إلى مجموعة بيانات تم جمعها علميًا.

Convergence

التقارب

يتم الوصول إلى حالة عندما تتغير قيم الخطأ قليلاً جدًا أو لا تتغير على الإطلاق مع كل تكرار. على سبيل المثال، يشير منحنى الخطأ التالي إلى تقارب عند حوالي 700 تكرار:



يتقارب (converges) النموذج عندما لا يؤدي التدريب الإضافي إلى تحسين النموذج.

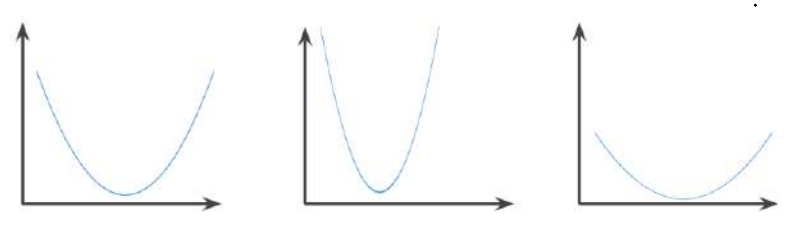
في التعلم العميق (deep learning)، تظل قيم الخطأفي بعض الأحيان ثابتة أو تقريبًا للعديد من التكرارات قبل أن تنخفضفي النهاية. خلال فترة طويلة من قيم الخطأ الثابتة، قد تحصل مؤقتًا على شعور زائف بالتقارب.

انظر أيضًا التوقف المبكر (early stopping).

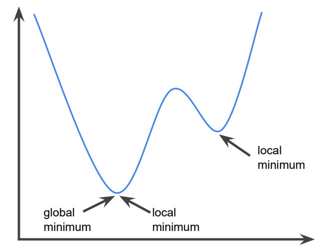
Convex function

دالة محدبة

دالة تكون فيها المنطقة فوق الرسم البياني للدالة عبارة عن مجموعة محدبة (convex set). تتشكل الدالة المحدبة النموذجية بشكل يشبه الحرف U. على سبيل المثال، كل ما يلي عبارة عن دوال محدبة:



في المقابل، الدالة التالية ليست محدبة. لاحظ كيف أن المنطقة الموجودة أعلى الرسم البياني ليست مجموعة محدبة:



تحتوي الدالة المحدبة بدقة (strictly convex function) على نقطة دنيا محلية واحدة، وهي أيضًا النقطة الدنيا العالمية. الدوال الكلاسيكية على شكل حرف U هي دوال محدبة بدقة. ومع ذلك، فإن بعض الدوال المحدبة (على سبيل المثال، الخطوط المستقيمة) ليست على شكل حرف U.

Convex optimization

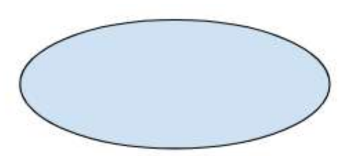
التحسين المحدب

عملية استخدام التقنيات الرياضية مثل التدرج الاشتقاقي (gradient descent) لإيجاد الحد الأدنى للدالة المحدبة (convex function). ركز قدر كبير من البحث في التعلم الآلي على صياغة مشاكل مختلفة كمشكلات تحسين محدبة وفي حل هذه المشكلات بشكل أكثر كفاءة.

convex set

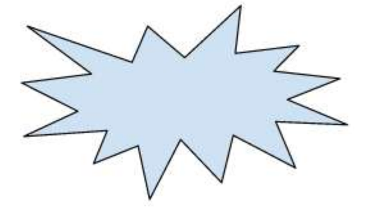
مجموعة محدبة

مجموعة فرعية من الفضاء الإقليدي بحيث يظل الخط المرسوم بين أي نقطتين في المجموعة الفرعية بالكامل داخل المجموعة الفرعية. على سبيل المثال، الشكلان التاليان عبارة عن مجموعات محدبة:





في المقابل، الشكلان التاليان ليسا مجموعات محدبة:





Convolution

التفاف

في الرياضيات، بشكل عرضي، مزيج من دالتين.في التعلم الآلي، يمزج الالتفاف بين المرشح التلافيفي (convolutional filter) ومصفوفة الإدخال من أجل تدريب الأوزان.

غالبًا ما يكون مصطلح "الالتفاف"في التعلم الآلي طريقة مختصرة للإشارة إلى العملية التلافيفية (convolutional operation) أو الطبقة التلافيفية (convolutional layer).

بدون التلافيف، سيتعين على خوارزمية التعلم الآلي أن تتعلم وزنًا منفصلاً لكل خليةفي موتر (tensor) كبير. على سبيل المثال، سيتم إجبار خوارزمية التعلم الآلي التي تتدرب على صور (2K x 2K)على إيجاد 4M أوزان منفصلة. بفضل التلافيف، يتعين على خوارزمية التعلم الآلي فقط العثور على أوزان لكل خليةفي المرشح (الفلتر) التلافيفي، مما يقلل بشكل كبير من الذاكرة اللازمة لتدريب النموذج. عند تطبيق المرشح التلافيفي، يتم نسخه ببساطة عبر الخلايا بحيث يتم ضرب كل منها بواسطة المرشح.

Convolutional filter

أحد الفاعلينفي عملية تلافيفية (convolutional operation). (العامل الآخر هو شريحة من مصفوفة الإدخال.) المرشح التلافيفي عبارة عن مصفوفة لها نفس رتبة مصفوفة الإدخال، ولكنها ذات شكل أصغر. على سبيل المثال، بالنظر إلى مصفوفة إدخال 28 × 28، يمكن أن يكون المرشح أي مصفوفة ثنائية الأبعاد أصغر من 28 × 28.

في المعالجة الفوتوغرافية، يتم عادةً تعيين جميع الخلايافي المرشح التلافيفي على نمط ثابت من الآحاد والأصفار.في التعلم الآلي، عادةً ما يتم زرع المرشحات التلافيفية بأرقام عشوائية ثم تقوم الشبكة بتدريب (trains) القيم المثالية.

Convolutional layer

طبقة تلافيفية

طبقة من شبكة عصبية عميقة (deep neural network) يمر فيها مرشح تلافيفي (convolutional filter) عبر مصفوفة إدخال. على سبيل المثال، ضعفي اعتبارك المرشح التلافيفي 3x3 التالي:

0	1	0
1	0	1
0	1	0

يوضح الرسم المتحرك التالي طبقة تلافيفية تتكون من 9 عمليات تلافيفية تتضمن مصفوفة إدخال 5 × 5. لاحظ أن كل عملية تلافيفية تعمل على شريحة 3x3مختلفة من مصفوفة الإدخال. تتكون المصفوفة الناتجة 3x3 (على اليمين) من نتائج 9 عمليات تلافيفية:

128	97	53	201	198
35	22	25	200	195
37	24	28	197	182
33	28	92	195	179
31	40	100	192	177

181		

Convolutional Neural Network (CNN)

الشبكة العصبية التلافيفية

شبكة عصبية (neural network) تكون فيها طبقة واحدة على الأقل طبقة تلافيفية. تتكون الشبكة العصبية التلافيفية النموذجية من مزيج من الطبقات التالية:

- طبقات تلافيفية (convolutional layers).
- طبقات التجميع (pooling layers).
- طبقات كثيفة (dense layers).

حققت الشبكات العصبية التلافيفية نجاحًا كبيرًا في أنواع معينة من المشكلات، مثل التعرف على الصور (image recognition).

Convolutional operation

عملية تلافيفية

العملية الحسابية التالية المكونة من خطوتين:

1. الضرب بالعناصر للمرشح التلافيفي (convolutional filter) وشريحة من مصفوفة الادخال. (شريحة مصفوفة الإدخال لها نفس رتبة وحجم المرشح التلافيفي.)
2. جمع كل القيم في مصفوفة حاصل الضرب الناتجة.

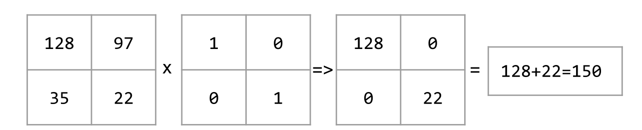
على سبيل المثال، ضع في اعتبارك مصفوفة الإدخال التالية 5 × 5:

128	97	53	201	198
35	22	25	200	195
37	24	28	197	182
33	28	92	195	179
31	40	100	192	177

تخيل الآن الفلتر التلافيفي 2x2 التالي:

1	0
0	1

تتضمن كل عملية تلافيفية شريحة مفردة 2x2 من مصفوفة الإدخال. على سبيل المثال، لنفترض أننا نستخدم شريحة 2x2 في أعلى يسار مصفوفة الإدخال. لذلك، تبدو عملية الالتفاف على هذه الشريحة كما يلي:



تتكون الطبقة التلافيفية (convolutional layer) من سلسلة من العمليات التلافيفية، تعمل كل منها على شريحة مختلفة من مصفوفة الإدخال.

Correlation

الارتباط

الارتباط هو نسبة التغاير بين متغيرين إلى منتج التباين (للمتغيرات). يأخذ قيمة بين 1+ و1_ . تعني القيمة القصوى على كلا الجانبين أنهما مرتبطان ارتباطًا وثيقًا ببعضهما البعض. تشير القيمة الصفرية إلى وجود ارتباط صفري ولكن ليس عدم الاعتماد.

Correlation matrix

مصفوفة الارتباط

مصفوفة الارتباط هي ببساطة جدول يعرض معاملات الارتباط (correlation coefficients) للمتغيرات المختلفة. توضح المصفوفة الارتباط بين جميع أزواج القيم الممكنة في الجدول. إنها أداة قوية لتلخيص مجموعة بيانات كبيرة ولتحديد وتصور الأنماط في البيانات المقدمة.

Corpus

المجموعة

المجموعة عبارة عن مجموعة من النصوص الأصلية أو الصوت المنظم في مجموعات البيانات. أصيلة هنا تعني النص المكتوب أو الصوتي الذي يتحدث به مواطن من اللغة أو اللهجة. يمكن أن تتكون المجموعة من كل شيء من الصحف والروايات والوصفات والبث الإذاعي إلى البرامج التلفزيونية والأفلام والتغريدات

Cost

الكلفة

مرادف للخطأ (loss).

Cost Function

دالة الكلفة

تستخدم دالة التكلفة لتحديد وقياس خطأ النموذج.

مرادف لدالة الخسارة(الخطأ) (loss function)

Co-training

التدريب المشترك

يعتبر نهج التعلم شبه الخاضع للإشراف (semi-supervised learning) مفيداً بشكل خاص عندما تكون جميع الشروط التالية صحيحة:

- نسبة الأمثلة غير المسماة (unlabeled examples) إلى الأمثلة المسماة (labeled examples)في مجموعة البيانات عالية.
- هذه مشكلة تصنيف (ثنائية (binary) أو متعددة الفئات (multi-class)).
- تحتوي مجموعة البيانات (dataset) على مجموعتين مختلفتين من الميزات التنبؤية المستقلة عن بعضها البعض والمتكاملة.

يعمل التدريب المشترك بشكل أساسي على تضخيم الإشارات المستقلة إلى إشارة أقوى. على سبيل المثال، ضعفي اعتبارك نموذج تصنيف (classification model) يصنف السيارات المستعملة الفردية على أنها جيدة أو سيئة. قد تركز مجموعة واحدة من الميزات التنبؤية على الخصائص الإجمالية مثل سنة السيارة ومنتجها وطرازها؛ قد تركز مجموعة أخرى من الميزات التنبؤية على سجل قيادة المالك السابق وتاريخ صيانة السيارة.

Covariance

التغاير

التغاير هو مقياس للتغير المشترك لمتغيرين عشوائيين. إنه مشابه للتباين (variance)، ولكن عندما يخبرك التباين بكيفية اختلاف متغير واحد، فإن التباين المشترك يخبرك بكيفية اختلاف متغيرين معًا.

Coverage bias

تحيز التغطية

انظر تحيز الاختيار (selection bias).

Crash blossom

جملة أو عبارة ذات معنى غامض. تمثل crash blossom مشكلة كبيرة في فهم اللغة الطبيعية. على سبيل المثال، العنوان Red Tape Holds Up Skyscraper هو crash blossom لأن نموذج NLU يمكن أن يفسر العنوان حرفيًا أو مجازيًا.

Critic

الناقد

مرادف لـ (Deep Q-Network).

Cross-entropy

الانتروبيا المتقاطعة

تعميم لخسارة السجل (Log Loss) على مشاكل التصنيف متعددة الفئات (multi-class classification problems). تحدد الانتروبيا الفرق بين توزيعين احتماليين. انظر أيضا الارتباك (perplexity).

Cross-validation

التحقق المتقاطع

آلية لتقدير مدى التعميم الجيد للنموذج (model) على البيانات الجديدة عن طريق اختبار النموذج مقابل مجموعة فرعية واحدة أو أكثر من البيانات غير المتداخلة التي تم حجبها من مجموعة التدريب (training set).

CUDA

CUDA عبارة عن منصة حوسبة متوازية وواجهة برمجة تطبيقات تسمح للبرنامج باستخدام أنواع معينة من وحدات معالجة الرسومات (GPU) لمعالجة الأغراض العامة، وهو نهج يسمى الحوسبة ذات الأغراض العامة على وحدات معالجة الرسومات.

D

D

Data analysis

تحليل البيانات

الحصول على فهم للبيانات من خلال النظرفي العينات والقياس والرسم. يمكن أن يكون تحليل البيانات مفيداً بشكل خاص عند استلام مجموعة البيانات لأول مرة، قبل أن يبني المرء النموذج (model) الأول. كما أنه مهمفي فهم التجارب وتصحيح الأخطاءفي النظام.

Data augmentation

زيادة البيانات

الزيادة المصطنعة لنطاق وعدد أمثلة التدريب (training) عن طريق تحويل الأمثلة الموجودة لإنشاء أمثلة إضافية. على سبيل المثال، افترض أن الصور هي إحدى ميزاتك (features)، لكن مجموعة البيانات الخاصة بك لا تحتوي على أمثلة صور كافية للنموذج لتعلم الارتباطات المفيدة. من الناحية المثالية، يمكنك إضافة ما يكفي من الصور المصنفة (labeled) إلى مجموعة البيانات الخاصة بك لتمكين نموذجك من التدريب بشكل صحيح. إذا لم يكن ذلك ممكنًا، يمكن أن تقوم زيادة البيانات بتدوير وتمديد وعكس كل صورة لإنتاج العديد من المتغيرات من الصورة الأصلية، مما قد ينتج عنه بيانات مصنفة كافية لتمكين التدريب الممتاز.

DataFrame

إطار البيانات

نوع بيانات شائع من Pandas لتمثيل مجموعات البياناتفي الذاكرة.

يعتبر DataFrame مماثلاً لجدول أو جدول بيانات. كل عمودفي DataFrame له اسم (رأس)، ويتم تحديد كل صف برقم فريد.

يتم تنظيم كل عمودفي DataFrame مثل مصفوفة ثنائية الأبعاد، باستثناء أنه يمكن تعيين نوع البيانات الخاص به لكل عمود.

Data Mining

التنقيب في البيانات

التنقيبفي البيانات هو دراسة لاستخراج المعلومات المفيدة من البيانات المهيكلة / غير المهيكلة المأخوذة من مصادر مختلفة. يتم ذلك عادة من أجل:

- التنقيب لأنماط متكررة.
- التنقيب للاشتراكات.
- التنقيب للارتباطات.
- التنقيب للمجموعات.
- التنقيب للتحليل التنبئي.

يتم التنقيب عن البيانات لأغراض مثل تحليل السوق، وتحديد نمط شراء العملاء، والتخطيط المالي، واكتشاف الاحتيال، وما إلى ذلك

Data Pipeline

مسار تدفق البيانات

مسار تدفق البيانات هو عملية أتمتة سير العمل لمهمة التعلم الآلي الكاملة. يمكن القيام بذلك عن طريق تمكين تحويل سلسلة من البيانات وربطها معًافي نموذج يمكن تحليله للحصول على المخرجات.

Dataset

مجموعة البيانات

مجموعة من البيانات الأولية، يتم تنظيمها بشكل شائع (ولكن ليس حصريًا) بأحد التنسيقات التالية:

- جدول بيانات (spreadsheet).
- ملف بتنسيق CSV (قيم مفصولة بفواصل comma-separated values).

Data Transformation

تحويل البيانات

تحويل البيانات هو عملية تحويل البيانات من نموذج إلى آخر. يتم ذلك عادةفي خطوة معالجة مسبقة.

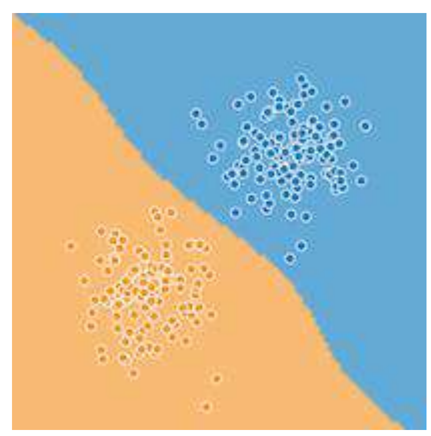
DBSCAN

إنها خوارزمية تجميع غير بارامترية قائمة على الكثافة (density-based clustering): بالنظر إلى مجموعة من النقاطفي مساحة ما، فإنها تجمع معًا النقاط التي يتم تجميعها معًا بشكل وثيق (نقاط مع العديد من الجيران القريبين)، ووضع علامة على أنها نقاط متطرفة تقع بمفردهافي المناطق منخفضة الكثافة (أقرب جيرانهم بعيدون جدًا). DBSCAN هي واحدة من أكثر خوارزميات التجميع شيوعًا وأكثرها اقتباسًا أيضًافي الأدبيات العلمية.

Decision boundary

حدود القرار

الفاصل بين الفئات (classes) التي تعلمها نموذج (model)في فئة ثنائية (binary class) أو مشاكل تصنيف متعددة الفئات (multi-class classification). على سبيل المثال،في الصورة التالية التي تمثل مشكلة تصنيف ثنائي، تكون حدود القرار هي الحد الفاصل بين الفئة البرتقالية والفئة الزرقاء:



Decision forest

غابة القرار

نموذج تم إنشاؤه من عدة أشجار قرار (decision trees). تقوم غابة القرار بعمل تنبؤ من خلال تجميع تنبؤات أشجار القرار الخاصة بها. تشمل الأنواع الشائعة من غابات القرار الغابات العشوائية (random forests) والأشجار المعززة بالتدرج (gradient boosted trees).

Decision Rules

قاعدة القرار

قاعدة القرار هي عبارة IF-THEN بسيطة تتكون من شرط (condition) وتوقع (prediction). على سبيل المثال: إذا أمطرت اليوم وإذا كانت أبريل (شرط)، فسوف تمطر غداً (التنبؤ). يمكن استخدام قاعدة قرار واحدة أو مجموعة من عدة قواعد لعمل تنبؤات.

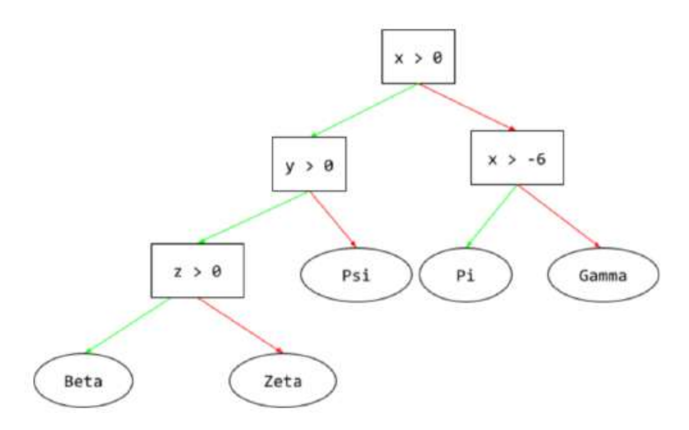
Decision threshold

مرادف لعتبة التصنيف (classification threshold).

Decision tree

شجرة القرار

نموذج تعلم خاضع للإشراف يتكون من مجموعة من الشروط (conditions) والأوراق (leaves) المنظمة بشكل هرمي. على سبيل المثال، ما يلي عبارة عن شجرة قرارات:



Deep model

نموذج عميق

شبكة عصبية (neural network) تحتوي على أكثر من طبقة مخفية (hidden layer).

يُطلق على النموذج العميق أيضًا اسم الشبكة العصبية العميقة (deep neural network).

على النقيض من النموذج الواسع (wide model).

Deep Learning

التعلم العميق

التعلم العميق هو مجموعة فرعية من التعلم الآلي (machine learning)، وهوفي الأساس شبكة عصبية ذات ثلاث طبقات أو أكثر. تحاول هذه الشبكات العصبية محاكاة سلوك الدماغ البشري– وإن كان بعيداً عن مطابقة قدرته – مما يسمح له "بالتعلم" من كميات كبيرة من البيانات. بينما لا يزال بإمكان الشبكة العصبية ذات الطبقة الواحدة إجراء تنبؤات تقريبية، يمكن أن تساعد الطبقات المخفية (hidden layer) الإضافيةفي تحسين الدقة وتحسينها.

Dependent Variable

المتغير التابع

المتغير التابع هو ما تقيسه والذي يتأثر بالمتغير (المتغيرات) المستقلة (independent variable) / الإدخال. يطلق عليه "تابع" لأنه "يعتمد" على المتغير المستقل. على سبيل المثال،

لنفترض أننا نريد توقع عادات التدخين لدى الأشخاص. ثم يدخن الشخص "نعم" أو "لا" هو المتغير التابع.

Decoder

مفكك الشفرة

بشكل عام، أي نظام تعلم آلي يتحول من تمثيل معالج أو كثيف أو داخلي إلى تمثيل أكثر خامًا أو متفرقة أو خارجيًا.

غالبًا ما تكون مفككات الشفرة أحد مكونات نموذج أكبر، حيث يتم إقرانهافي كثير من الأحيان بمشفر (encoder).

في مهام التسلسل إلى التسلسل (sequence-to-sequence tasks)، يبدأ مفكك الشفرة بالحالة الداخلية التي يولدها المشفر للتنبؤ بالتسلسل التالي.

راجع المحول (Transformer) لتعريف مفكك الشفرة داخل بنية المحولات.

Deep Neural Network

الشبكة العصبية العميقة

مرادف للنموذج العميق (deep model).

Deep Q-Network (DQN)

شبكة Q العميقة (DQN)

في (Q-Learning)، شبكة عصبية (neural network) عميقة تتنبأ بدوال Q (Q-functions).

الناقد (Critic) هو مرادف لـ(Deep Q-Network).

Denoising

ازالة الضوضاء

نهج شائع للتعلم تحت الإشراف الذاتي (self-supervised learning) حيث:

- يتم إضافة الضوضاء (Noise) بشكل مصطنع إلى مجموعة البيانات.
- يحاول النموذج (model) إزالة الضوضاء.

يتيح تقليل الضوضاء التعلم من الأمثلة غير المسماة (unlabeled examples). تعمل مجموعة البيانات (dataset) الأصلية كهدف أو تسمية (label) والبيانات الصاخبة (noisy data) كإدخال.

تستخدم بعض نماذج اللغة المقنعة (masked language models) تقليل الضوضاء على النحو التالي:

- يتم إضافة الضوضاء بشكل مصطنع إلى جملة غير مسماة بإخفاء بعض الرموز.
- يحاول النموذج توقع الرموز المميزة (الوحدة اللفظية) (tokens) الأصلية.

Dense feature

ميزة كثيفة

ميزة (feature) تكون معظم القيم فيها أو كلها غير صفرية، عادةً ما تكون موترًا (Tensor) لقيم الفاصلة العائمة. على سبيل المثال، عنصر الموتر التالي المكون من 10 عناصر كثيف لأن 9 من قيمه ليست صفرية:

8	3	7	5	2	4	0	4	9	6

على النقيض من ميزة متفرقة (sparse feature).

Dense layer

طبقة كثيفة

مرادف للطبقة المتصلة بالكامل (fully connected layer).

Dense Network

الشبكة الكثيفة

الشبكة الكثيفة هي شبكة يكون فيها عدد الروابط لكل عقدة قريبًا من العدد الأقصى للعقد. ترتبط كل عقدة بجميع العقد الأخرى تقريبًا. تسمى الحالة الإجمالية المتصلة التي ترتبط فيها كل عقدة ببعضها البعض بشبكة متصلة تمامًا (fully connected network).

Depth

عمق

مجموع ما يلي في الشبكة العصبية (neural network):

- عدد الطبقات المخفية (hidden layers).

- عدد طبقات الإخراج (output layers)، والذي يكون عادةً 1.
- عدد أي طبقات تضمين (embedding layers).

على سبيل المثال، يبلغ عمق الشبكة العصبية المكونة من خمس طبقات مخفية وطبقة إخراج واحدة 6.

لاحظ أن طبقة الإدخال (input layer) لا تؤثر على العمق.

Depthwise separable convolutional neural network (sepCNN)

شبكة عصبية تلافيفية قابلة للفصل بعمق (sepCNN)

بُنية شبكة عصبية تلافيفية (convolutional neural network) مبنية على (Inception)، ولكن حيث يتم استبدال وحدات Inception بتلافيفات قابلة للفصل بعمق. يُعرف أيضًا باسم(Xception).

التفاف قابل للفصل بعمق (يُختصر أيضًا باسم التفاف قابل للفصل separable convolution) يعمل على تحويل التفاف قياسي ثلاثي الأبعاد إلى عمليتي التفاف منفصلتين أكثر كفاءة من الناحية الحسابية: أولاً، التفاف عميق بعمق (1 X n X n) 1، ثم الثانية، الالتفاف النقطي، بطول وعرض (n X 1 X 1) 1.

Derived label

التسمية المشتقة

مرادف لتسمية الوكيل (proxy label).

Device

جهاز

فئة من الأجهزة يمكنها تشغيل جلسة TensorFlow، بمافي ذلك CPUs وGPUs و(TPUs).

Dimensions reduction

تقليل الابعاد

تقليل عدد الأبعاد المستخدمة لتمثيل ميزة معينةفي متجه المعالم، عادةً عن طريق التحويل إلى متجه تضمين (embedding vector).

Dimensions

الابعاد

مصطلح مثقل به أي من التعريفات التالية:

- عدد مستويات الإحداثيات في الموتر (Tensor). فمثلا:
  - العددية له أبعاد صفرية؛ على سبيل المثال، ["Hello"].
  - المتجه له بعد واحد؛ على سبيل المثال، [3 ، 5 ، 7 ، 11].
  - المصفوفة لها بعدين؛ على سبيل المثال، [[2 ، 4 ، 18] ، [5 ، 7 ، 14]].
- يمكنك تحديد خلية معينة بشكل فريدفي متجه أحادي البعد بإحداثيات واحدة؛ أنت بحاجة إلى إحداثيين لتحديد خلية معينة بشكل فريدفي مصفوفة ثنائية الأبعاد.
- عدد الإدخالات في متجه الميزة (feature vector).
- عدد العناصرفي طبقة التضمين (embedding layer).

Discrete feature

ميزة منفصلة

ميزة (feature) ذات مجموعة محدودة من القيم الممكنة. على سبيل المثال، الميزة التي قد تكون قيمها حيوانية أو نباتية أو معدنية هي ميزة منفصلة discrete (أو فئوية categorical).

على النقيض من الميزة المستمرة (continuous feature).

Discriminative model

نموذج تمييزي

نموذج (model) يتنبأ بالتسميات (labels) من مجموعة من المعالم (features) أو أكثر. بشكل أكثر رسمية، تحدد النماذج التمييزية الاحتمال الشرطي لمخرجات بالنظر إلى الميزات والأوزان (weights)؛ هذا هو:

```
p(output | features, weights)
```

على سبيل المثال، النموذج الذي يتنبأ بما إذا كانت رسالة البريد الإلكتروني هي بريد عشوائي من الميزات والأوزان هو نموذج تمييزي.

الغالبية العظمى من نماذج التعلم تحت الإشراف، بمافي ذلك نماذج التصنيف والانحدار، هي نماذج تمييزية.

على النقيض من النموذج التوليدي (generative model).

Discriminator

المميز

نظام يحدد ما إذا كانت الأمثلة حقيقية أم مزيفة.

بدلاً من ذلك، النظام الفرعي داخل شبكة الخصومة التوليدية (generative adversarial network (GAN)) يحدد ما إذا كانت الأمثلة التي تم إنشاؤها بواسطة المولد (generator) حقيقية أم مزيفة.

Divisive clustering

المجموعات الانقسامية

انظر التكتلات الهرمية (hierarchical clustering).

Dot Product

الضرب النقطي

الضرب النقطي هو طريقة لضرب متجهين متساويين الطول معًا. من الناحية المفاهيمية، هو مجموع حاصل ضرب العناصر المقابلةفي المتجهين. تتضمن الأسماء الأخرى لنفس العملية ما يلي: الضرب القياسي (Scalar product)، لأن النتيجة تنتج رقمًا قياسيًا (Scalar) واحداً.

Downsampling

الاختزال

مصطلح الاختزال (Downsampling) أن يعني أيًا مما يلي:

- تقليل كمية المعلوماتفي الميزة (feature) لتدريب (train) نموذج بشكل أكثر كفاءة. على سبيل المثال، قبل تدريب نموذج التعرف على الصور، اختزل الصور عالية الدقة إلى تنسيق أقل دقة.
- التدريب على نسبة منخفضة بشكل غير متناسب من امثله الفئة (class) ذات التمثيل الزائد من أجل تحسين التدريب النموذجيفي الفئات ذات التمثيل المنخفض. على سبيل المثال،في مجموعة بيانات غير متوازنةفي الفئة، تميل النماذج إلى تعلم الكثير عن فئة الأغلبية (majority class) وليس ما يكفي عن فئة الأقلية (minority class). يساعد الاختزالفي تحقيق التوازن بين مقدار التدريب على فئتي الأغلبية والأقلية.

DQN

اختصار لـشبكة Q العميقة (Deep Q-Network).

Dropout

التسرب

التسرب (الحذف العشوائي) هو تقنية يتم فيها تجاهل الخلايا العصبية المختارة عشوائيًا أثناء التدريب. هم "تسربوا" بشكل عشوائي. هذا يعني أن مساهمتهافي تنشيط الخلايا العصبية المصب يتم إزالتها مؤقتًا عند المرور الأمامي (forward pass)، ولا يتم تطبيق أي تحديثات للوزن على الخلايا العصبية الموجودةفي المرور الخلفي (backward pass).

Dropout regularization

تنظيم التسرب

شكل من أشكال التنظيم (regularization) مفيدفي تدريب الشبكات العصبية (neural networks). يزيل تنظيم التسرب تحديداً عشوائيًا لعدد ثابت من الوحداتفي طبقة الشبكة لخطوة واحدة من التدرج الاشتقاقي (gradient descent). وكلما زاد عدد الوحدات المسربة، كان التنظيم أقوى. هذا مشابه لتدريب الشبكة لمحاكاة مجموعة كبيرة بشكل كبير من الشبكات الأصغر.

Dummy Variable

المتغير الوهمي

المتغير الوهمي هو اسم آخر للمتغير المنطقي. مثال على المتغير الوهمي هو أنه يأخذ القيمة 0 أو 1. 0 يعني أن القيمة صحيحة (أي العمر <25) ويعني 1 أن القيمة خاطئة (أي العمر >= 25).

Dynamic

الديناميكي

شيء يتم القيام به بشكل متكرر أو مستمر. المصطلحان ديناميكي (dynamic) واونلاين (online) مترادفانفي التعلم الآلي. فيما يلي الاستخدامات الشائعة للديناميكية والاونلاينفي التعلم الآلي:

- النموذج الديناميكي dynamic model (أو النموذج الاونلاين online model) هو نموذج يتم إعادة تدريبه بشكل متكرر أو مستمر.
- التدريب الديناميكي Dynamic training (أو التدريب الاونلاين online training) هو عملية التدريب بشكل متكرر أو مستمر.
- الاستدلال الديناميكي Dynamic inference (أو الاستدلال الاونلاين online inference) هو عملية توليد التنبؤات عند الطلب.

Dynamic model

نموذج ديناميكي

نموذج (model) يتم إعادة تدريبه بشكل متكرر (ربما بشكل مستمر). النموذج الديناميكي هو "المتعلم مدى الحياة lifelong learner " الذي يتكيف باستمرار مع البيانات المتطورة. يُعرف النموذج الديناميكي أيضًا بالنموذج الاونلاين (online model).

على النقيض من النموذج الثابت (static model).

E

E

Edge detection

كشف الحواف

يعد اكتشاف الحواف تقنية لمعالجة الصور لإيجاد حدود الكائنات داخل الصور. إنه يعمل عن طريق الكشف عن الانقطاعات في السطوع. يُستخدم اكتشاف الحواف لتجزئة الصور واستخراج البيانات في مجالات مثل معالجة الصور والرؤية الحاسوبية (computer vision) ورؤية الآلة.

EigenValues

القيم الذاتية

القيم الذاتية هي مجموعة خاصة من القيم العددية المرتبطة بمجموعة المعادلات الخطية على الأرجح في معادلات المصفوفة. يُطلق على المتجهات الذاتية (eigenvectors) أيضًا جذور مميزة. إنه متجه غير صفري يمكن تغييره على الأكثر من خلال عامله القياسي بعد تطبيق التحويلات الخطية.

EigenVector

المتجهات الذاتية

المتجهات الذاتية هي مجموعة خاصة من المتجهات المرتبطة بنظام خطي من المعادلات (أي معادلة مصفوفة) والتي تُعرف أحيانًا أيضًا باسم المتجهات المميزة أو المتجهات المناسبة أو المتجهات الكامنة. يتم استخدام المتجهات الذاتية لجعل التحويل الخطي مفهومًا. فكر في المتجهات الذاتية على أنها تمدد / ضغط مخطط خطي X-Y دون تغيير اتجاهها.

Early stopping

التوقف المبكر

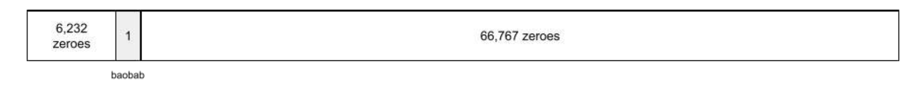
طريقة للتنظيم (regularization) تتضمن إنهاء التدريب قبل أن ينتهي خطأ التدريب في التناقص. في التوقف المبكر، تتوقف عمدًا عن تدريب النموذج عندما يبدأ الخطأ في مجموعة بيانات التحقق (validation dataset) في الزيادة؛ أي عندما يسوء أداء التعميم (generalization).

Embedding layer

طبقة التضمين

طبقة مخفية (hidden layer) خاصة تتدرب على ميزة فئوية (categorical feature) عالية الأبعاد لتتعلم تدريجياً متجه تضمين بُعد أقل. تمكّن طبقة التضمين الشبكة العصبية من التدريب بكفاءة أكبر بكثير من التدريب فقط على الميزة الفئوية عالية الأبعاد.

على سبيل المثال، تدعم الأرض حاليًا حوالي 73000 نوع من الأشجار. افترض أن أنواع الأشجار هي ميزةفي نموذجك، لذا تتضمن طبقة الإدخالفي نموذجك متجهًا واحداً ساخناً (one-hot vector) بطول 73000 عنصر. على سبيل المثال، ربما يتم تمثيل شجر الباوباب بشيء مثل هذا:



مصفوفة مكونة من 73000 عنصر طويلة جداً. إذا لم تقم بإضافة طبقة التضمين إلى النموذج، فسيستغرق التدريب وقتًا طويلاً للغاية بسبب مضاعفة 72،999 صفرًا. ربما تختار طبقة التضمين لتتكون من 12 بُعدًا. وبالتالي، ستتعلم طبقة التضمين تدريجياً متجه تضمين جديد لكل نوع من أنواع الأشجار.

في حالات معينة، تعد التجزئة (hashing) بديلاً معقولاً لطبقة التضمين.

Embedding vector

متجه التضمين

بشكل عام، مجموعة من أرقام الفاصلة العائمة مأخوذة من أي طبقة مخفية (hidden layer) تصف المدخلات إلى تلك الطبقة المخفية. غالبًا ما يكون متجه التضمين هو مصفوفة أرقام الفاصلة العائمة التي يتم تدريبهافي طبقة التضمين. على سبيل المثال، افترض أن طبقة التضمين يجب أن تتعلم متجه تضمين لكل نوع من أنواع الأشجار البالغ عددها 73000 نوع على الأرض. ربما تكون المصفوفة التالية هي متجه التضمين لشجرة الباوباب:

0.819245	0.539102	0.391284	0.181923	0.519247	0.912043	0.291529	0.519284	0.710153	0.109152	0.851024	0.381528

متجه التضمين ليس مجموعة من الأرقام العشوائية. تحدد طبقة التضمين هذه القيم من خلال التدريب، على غرار الطريقة التي تتعلم بها الشبكة العصبية الأوزان الأخرى أثناء التدريب. كل عنصر من عناصر المصفوفة هو تصنيف على طول بعض خصائص أنواع الأشجار. ما العنصر الذي يمثل خاصية أي نوع من الأشجار؟ من الصعب جداً على البشر تحديد ذلك.

الجزء الرائع رياضيًا من متجه التضمين هو أن العناصر المتشابهة لها مجموعات متشابهة من أرقام الفاصلة العائمة. على سبيل المثال، أنواع الأشجار المتشابهة لديها مجموعة من أرقام الفاصلة العائمة أكثر تشابهًا من أنواع الأشجار غير المتشابهة. الخشب الأحمر والسيكويا من

أنواع الأشجار ذات الصلة، لذلك سيكون لديهم مجموعة أكثر تشابهًا من الأرقام التي تشير إلى العائمة من الأخشاب الحمراء ونخيل جوز الهند. ستتغير الأرقام الموجودةفي متجه التضمينفي كل مرة تقوم فيها بإعادة تدريب النموذج، حتى إذا قمت بإعادة تدريب النموذج بمدخلات متطابقة.

Emotion recognition

التعرف على المشاعر

التعرف على المشاعر هو عملية التعرف على المشاعر البشرية. يختلف الناس بشكل كبيرفي دقتهم في التعرف على مشاعر الآخرين. يعد استخدام التكنولوجيا لمساعدة الأشخاصفي التعرف على المشاعر مجال بحث حديث العهد نسبيًا. بشكل عام، تعمل التقنية بشكل أفضل إذا كانت تستخدم طرائق متعددةفي السياق.

Empirical risk minimization (ERM)

تقليل المخاطر التجريبية

إنه مبدأفي نظرية التعلم الإحصائي الذي يعرف عائلة خوارزميات التعلم وتستخدم لإعطاء حدود نظرية على أدائها.

Encoder

المشفر

بشكل عام، أي نظام تعلم آلي (machine learning) يتحول من تمثيل خام أو متناثر أو خارجي إلى تمثيل داخلي أكثر معالجة أو كثافة أو أكثر.

غالبًا ما تكون المشفرات أحد مكونات نموذج أكبر، حيث يتم إقرانهافي كثير من الأحيان مع مفككات الشفرة (decoders). تقوم بعض المحولات (transformers) بإقران المشفرات مع مفككات الشفرة، على الرغم من أن المحولات الأخرى تستخدم فقط المشفر أو مفكك الشفرة فقط.

تستخدم بعض الأنظمة إخراج المشفر كمدخل إلى شبكة تصنيف أو انحدار.

في مهام التسلسل إلى التسلسل (sequence-to-sequence tasks)، يأخذ المشفر تسلسل إدخال ويعيد حالة داخلية (متجه). بعد ذلك، يستخدم مفكك الشفرة تلك الحالة الداخلية للتنبؤ بالتسلسل التالي.

راجع المحولات (Transformer) لتعريف المشفرفي بُنية المحولات.

Ensemble

التجمع

مجموعة من النماذج (models) المدربة بشكل مستقل والتي يتم حساب متوسط أو تجميع تنبؤاتها.في كثير من الحالات، تنتج المجموعة تنبؤات أفضل من نموذج واحد. على سبيل المثال، الغابة العشوائية (random forest) هي مجموعة مبنية من عدة أشجار قرار (description trees). لاحظ أنه ليست كل غابات القرار (decision forests) عبارة عن تجمعات (Ensembles).

Entropy

الانتروبيا

في نظرية المعلومات (information theory)، وصف لمدى عدم إمكانية التنبؤ بتوزيع الاحتمالات. بدلاً من ذلك، يتم تعريف الانتروبيا أيضًا على أنها مقدار المعلومات التي يحتوي عليها كل مثال. يحتوي التوزيع على أعلى إنتروبيا ممكنة عندما تكون جميع قيم المتغير العشوائي متساويةفي الاحتمال.

إنتروبيا مجموعة ذات قيمتين محتملتين "0" و "1" (على سبيل المثال، التسمياتفي مشكلة تصنيف ثنائي (binary classification)) لها الصيغة التالية:

```
H = -p log p - q log q = -p log p - (1-p) * log (1-p)
```

حيث:

- H هو الانتروبيا.
- p هو جزء من الأمثلة "1".
- q هو جزء من "0" أمثلة. لاحظ أن q = (1 - p)
- log بشكل عام هو log_2.في هذه الحالة، تكون وحدة الإنتروبيا قليلة.

على سبيل المثال، افترض ما يلي:

- يحتوي 100 مثال على القيمة "1"
- 300 مثال تحتوي على القيمة "0"

لذلك، فإن قيمة الانتروبيا هي:

- p = 0.25
- q = 0.75

- H = (-0.25)log2(0.25) - (0.75)log2(0.75) = 0.81 بت لكل مثال

المجموعة المتوازنة تمامًا (على سبيل المثال، 200 "0" ثانية و200 "1" ثانية) سيكون لها إنتروبيا 1.0 بت لكل مثال. عندما تصبح المجموعة غير متوازنة (imbalanced) أكثر، تتحرك الانتروبيا نحو 0.0.

في أشجار القرار (decision trees)، تساعد الأنتروبيافي تكوين اكتساب المعلومات (information gain) لمساعدة المُقسِّم (splitter) على تحديد الشروط (conditions) أثناء نمو شجرة قرار التصنيف.

قارن الانتروبيا بـ:

- gini impurity.
- دالة خطأ cross-entropy.

غالبًا ما يُطلق على الانتروبيا اسم إنتروبيا شانون (Shannon's entropy).

Environment

البيئة

في التعلم المعزز (reinforcement learning)، العالم الذي يحتوي على الوكيل (agent) ويسمح له بمراقبة حالة (state) هذا العالم. على سبيل المثال، يمكن أن يكون العالم الممثل لعبة مثل الشطرنج، أو عالمًا ماديًا مثل المتاهة. عندما يطبق الوكيل إجراءً (action) على البيئة، فإن البيئة تنتقل بين الحالات.

Episode

الحلقة

في التعلم المعزز (reinforcement learning)، كل من المحاولات المتكررة من قبل الوكيل لتعلم بيئة.

Epoch

الحقبة (الفترة)

تمريرة تدريب كاملة على مجموعة التدريب بأكملها بحيث تمت معالجة كل مثال مرة واحدة.

تمثل الحقبة تكرارات التدريب بتقسيم N على حجم الدفعة (batch size)، حيث N هو العدد الإجمالي للأمثلة.

على سبيل المثال، افترض ما يلي:

- تتكون مجموعة البيانات من 1000 مثال.
- حجم الدُفعة هو 50 مثالاً.

لذلك، تتطلب الحقبة الواحدة 20 تكرارًا:

1 epoch = (N/batch size) = (1,000 / 50) = 20 iterations

Epsilon greedy policy

سياسة الجشع إبسيلون

في التعلم المعزز (reinforcement learning)، وهي سياسة إما تتبع سياسة عشوائية (random policy) مع احتمال إبسيلون أو سياسة الجشع (greedy policy) خلاف ذلك. على سبيل المثال، إذا كانت قيمة ابسليون تساوي 0.9، فإن السياسة تتبع سياسة عشوائية بنسبة 90٪ من الوقت وسياسة جشع بنسبة 10٪ من الوقت.

على مدار الحلقات المتتالية، تقلل الخوارزمية من قيمة إبسيلون من أجل التحول من اتباع سياسة عشوائية إلى اتباع سياسة الجشع. من خلال تغيير السياسة، يستكشف الوكيل أولاً بشكل عشوائي البيئة ثم يستغل نتائج الاستكشاف العشوائي بجشع.

Example

مثال

قيم صف واحد من المعالم (features) وربما تسمية (label). تنقسم الأمثلةفي التعلم الخاضع للإشراف (supervised learning) إلى فئتين عامتين:

- يتكون المثال المسمى (labeled example) من واحد أو أكثر من المعالم والتسمية. تستخدم الأمثلة المصنفة أثناء التدريب.
- يتكون المثال غير المصنف (unlabeled example) من ميزة واحدة أو أكثر ولكن بدون تسمية. تستخدم الأمثلة غير المصنفة أثناء الاستدلال.

على سبيل المثال، افترض أنك تقوم بتدريب نموذج لتحديد تأثير الظروف الجوية على درجات اختبار الطالب. فيما يلي ثلاثة أمثلة معنونة:

Features			Label
Temperature	Humidity	Pressure	Test score
15	47	998	Good
19	34	1020	Excellent
18	92	1012	Poor

فيما يلي ثلاثة أمثلة غير مسماة:

Temperature	Humidity	Pressure
12	62	1014
21	47	1017
19	41	1021

عادةً ما يكون صف مجموعة البيانات هو المصدر الأولي لمثال. أي أن المثال يتكون عادةً من مجموعة فرعية من الأعمدةفي مجموعة البيانات. علاوة على ذلك، يمكن أن تتضمن الميزات الموجودةفي أحد الأمثلة أيضًا ميزات تركيبية (synthetic features)، مثل تقاطعات الميزات (feature crosses).

Experience replay

إعادة التجربة

في التعلم المعزز (reinforcement learning)، تستخدم تقنية (DQN) لتقليل الارتباطات الزمنيةفي بيانات التدريب. يخزن الوكيل (agent) انتقالات الحالةفي مخزن مؤقت للإعادة (replay buffer)، ثم يأخذ عينات من الانتقالات من المخزن المؤقت للإعادة لإنشاء بيانات التدريب.

Experimenter's bias

تحيز المجرب

انظر تحيز التأكيد (confirmation bias).

Exploding gradient problem

مشكلة انفجار الانحدار

الميل إلى التدرجات (gradients)في الشبكات العصبية العميقة (deep neural networks) (خاصة الشبكات العصبية المتكررة (recurrent neural networks)) لتصبح شديدة الانحدار (عالية) بشكل مدهش. غالبًا ما تسبب التدرجات الحادة تحديثات كبيرة جدًا لأوزان (weights) كل عقدة (node)في شبكة عصبية عميقة.

تصبح النماذج التي تعاني من مشكلة انفجار التدرج صعبة أو مستحيلة التدريب. يمكن أن يؤدي قص التدرج (Gradient clipping) إلى تخفيف هذه المشكلة.

قارن مع مشكلة تلاشي التدرج (vanishing gradient problem).

exploratory data analysis (EDA)

تحليل البيانات الاستكشافية (EDA)

تحليل البيانات الاستكشافية (EDA) هي مرحلة تستخدم لخط أنابيب علوم البيانات حيث يكون التركيز على فهم رؤى البيانات من خلال الرسم أو عن طريق التحليل الإحصائي

خطوات المشاركةفي تحليل البيانات الاستكشافية (EDA) هي:

1. تعريف المتغيرفي هذه الخطوة، نحدد نوع البيانات وفئة المتغيرات.
2. تحليل أحادي المتغير (Univariate analysis).
3. تحليل متعدد المتغيرات (Multivariate analysis).

Evaluation Metrics

مقاييس التقييم

الغرض من مقياس التقييم هو قياس جودة نموذج التعلم الإحصائي / الآلي. على سبيل المثال، فيما يلي بعض مقاييس التقييم:

1. AUC
2. ROC score
3. F-Score
4. Log-Loss

F

F

False negative (FN)

السلبي الخاطئ (FN)

مثال يتنبأ فيه النموذج بطريق الخطأ بالفئة السلبية (negative class). على سبيل المثال، يتنبأ النموذج بأن رسالة بريد إلكتروني معينة ليست بريداً عشوائيًا (الفئة السلبية)، ولكن رسالة البريد الإلكتروني هذه هيفي الواقع بريد عشوائي.

False negative rate

المعدل السلبي الخاطئ

نسبة الأمثلة الإيجابية الفعلية التي تنبأ بها النموذج عن طريق الخطأ بالفئة السلبية(negative class). الصيغة التالية تحسب المعدل السلبي الخاطئ:

$$\text{false negative rate} = \frac{\text{false negatives}}{\text{false negatives} + \text{true positives}}$$

False positive (FP)

الإيجابي الخاطئ (FP)

مثال يتنبأ فيه النموذج عن طريق الخطأ بالفئة الإيجابية (positive class). على سبيل المثال، يتنبأ النموذج بأن رسالة بريد إلكتروني معينة هي بريد عشوائي (الفئة الإيجابية)، لكن رسالة البريد الإلكتروني هذه ليستفي الواقع بريداً عشوائيًا.

False positive rate (FPR)

المعدل الإيجابي الخاطئ (FPR)

نسبة الأمثلة السلبية الفعلية التي تنبأ بها النموذج عن طريق الخطأ بالفئة الإيجابية. الصيغة التالية تحسب المعدل الإيجابي الخاطئ:

$$\text{false positive rate} = \frac{\text{false positives}}{\text{false positives} + \text{true negatives}}$$

المعدل الإيجابي الخاطئ هو المحور xفي منحنى ROC (ROC curve).

Feature

الميزة

متغير إدخال لنموذج التعلم الآلي. يتكون المثال (example) من ميزة واحدة أو أكثر. على سبيل المثال، افترض أنك تقوم بتدريب نموذج لتحديد تأثير الظروف الجوية على درجات اختبار الطالب. يوضح الجدول التالي ثلاثة أمثلة، يحتوي كل منها على ثلاث ميزات وتسمية واحدة:

Features			Label
Temperature	Humidity	Pressure	Test score
15	47	998	92
19	34	1020	84
18	92	1012	87

على النقيض من التسمية (label).

Feature cross

الميزة المتقاطعة

ميزة تركيبية (synthetic feature) تتشكل من خلال "تقاطع" السمات الفئوية (categorical) أو المجمعة (bucketed).

على سبيل المثال، ضعفي اعتبارك نموذج "التنبؤ بالحالة المزاجية" الذي يمثل درجة الحرارةفي أحد المجموعات الأربعة التالية:

- freezing
- chilly
- temperate
- warm

ويمثل سرعة الرياحفي أحد المجموعات التالية:

- still
- light
- windy

بدون ميزات متقاطعة، يتدرب النموذج الخطي بشكل مستقل على كل من المجموعات السبع المختلفة السابقة. لذلك، يتدرب النموذج، على سبيل المثال، على freezing بشكل مستقل عن التدريب، على سبيل المثال، windy.

بدلاً من ذلك، يمكنك إنشاء ميزة متقاطعة لدرجة الحرارة وسرعة الرياح. ستحتوي هذه الميزة التركيبية على القيم المحتملة الـ 12 التالية:

- freezing-still

- freezing-light
- freezing-windy
- chilly-still
- chilly-light
- chilly-windy
- temperate-still
- temperate-light
- temperate-windy
- warm-still
- warm-light
- warm-windy

بفضل الميزات المتقاطعة، يمكن للنموذج معرفة الاختلافات المزاجية بين يوم freezing-windy ويوم freezing-still.

إذا قمت بإنشاء ميزة تركيبية من ميزتين يحتوي كل منهما على الكثير من المجموعات المختلفة، فستحتوي الميزة المتقاطعة الناتج على عدد كبير من التركيبات الممكنة. على سبيل المثال، إذا كانت إحدى الميزات تحتوي على 1000 مجموعة والميزة الأخرى بها 2000 مجموعة، فإن الميزة المتقاطعة الناتج يحتوي على 2000000 مجموعة.

بشكل رسمي، التقاطع (cross) هو ضرب ديكارتي (Cartesian product).

تُستخدم الميزات المتقاطعةفي الغالب مع النماذج الخطية ونادرًا ما تُستخدم مع الشبكات العصبية.

Feature engineering

هندسة الميزات

عملية تتضمن الخطوات التالية:

1. تحديد الميزات (features) التي قد تكون مفيدةفي تدريب النموذج.
2. تحويل البيانات الأولية من مجموعة البيانات إلى إصدارات فعالة من تلك الميزات.

على سبيل المثال، قد تحدد أن درجة الحرارة قد تكون ميزة مفيدة. بعد ذلك، يمكنك تجربة استخدام (bucketing)لتحسين ما يمكن أن يتعلمه النموذج من نطاقات درجات الحرارة المختلفة.

تسمى هندسة الميزات أحيانًا استخراج الميزات (feature extraction).

Feature extraction

استخراج الميزة

مصطلح مثقل به أي من التعريفين التاليين:

- استرداد تمثيلات الميزات الوسيطة المحسوبة بواسطة نموذج غير خاضع للإشراف (unsupervised) أو تم اختباره مسبقًا (على سبيل المثال، قيم الطبقة المخفية (hidden layer)في شبكة عصبية) لاستخدامهافي نموذج آخر كمدخلات.
- مرادف لهندسة الميزات (feature engineering).

Feature importance

أهمية الميزات

مرادف لأهمية المتغيرات (variable importance).

Feature Map

خريطة الميزات

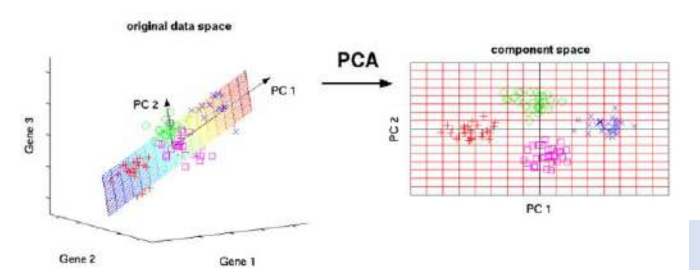
خريطة الميزات هي مصفوفة ثنائية الأبعاد من الخلايا العصبية. تستقبل الطبقة التلافيفية (convolutional layer) كتلة من خرائط ميزات الإدخال وتقوم بإنشاء كتلة من خرائط ميزات المخرجات.

Feature Reduction

تقليل الميزات

تقليل الميزات هو عملية تقليل عدد الميزات للعمل على مهمة حسابية مكثفة دون فقدان الكثير من المعلومات.

تعد تقنية تحليل المكونات الرئيسية (PCA) واحدة من أكثر تقنيات تقليل الميزات شيوعًا، حيث نقوم بدمج المتغيرات المترابطة لتقليل الميزات.



Feature Selection

اختيار الميزات

اختيار الميزة هو عملية تحديد الميزات ذات الصلة من مجموعة بيانات لإنشاء نموذج التعلم الآلي.

Feature set

مجموعة الميزات

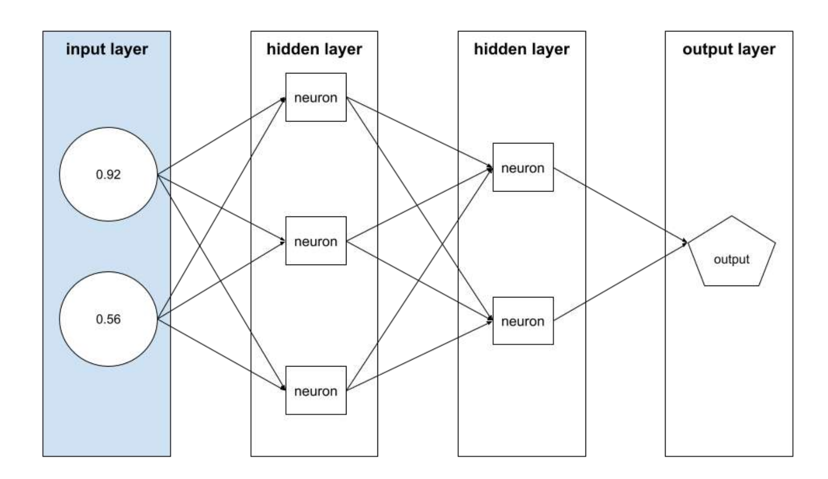
مجموعة الميزات (features) التي يتم تدريب نموذج (model) التعلم الآلي عليها. على سبيل المثال، قد يشتمل الرمز البريدي وحجم العقار وحالة العقار على مجموعة ميزات بسيطة لنموذج يتنبأ بأسعار المساكن.

Feature vector

متجه الميزات

مصفوفة قيم السمة (feature) تشتمل على مثال (example). يتم إدخال متجه الميزة أثناء التدريب (training) وأثناء الاستدلال (inference). على سبيل المثال، قد يكون متجه الميزة لنموذج بميزتين منفصلتين:

```
[0.92, 0.56]
```



يوفر كل مثال قيمًا مختلفة لمتجه الميزة، لذلك يمكن أن يكون متجه الميزة للمثال التالي شيئًا مثل:

```
[0.73, 0.49]
```

تحدد هندسة الميزات (Feature engineering) كيفية تمثيل الميزاتفي متجه الميزات. على سبيل المثال، قد يتم تمثيل ميزة فئوية ثنائية بخمس قيم محتملة بترميز واحد ساخن (one-hot encoding).في هذه الحالة، سيتكون جزء متجه الميزة لمثال معين من أربعة أصفار و1.0 واحد في الموضع الثالث، على النحو التالي:

```
[0.0, 0.0, 1.0, 0.0, 0.0]
```

كمثال آخر، افترض أن نموذجك يتكون من ثلاث ميزات:

- ميزة فئوية ثنائية مع خمس قيم محتملة ممثلة بترميز واحد ساخن؛ على سبيل المثال: [0.0 ، 0.0 ، 0.0 ، 1.0 ، 0.0]
- ميزة فئوية ثنائية أخرى بثلاث قيم محتملة ممثلة بترميز واحد ساخن؛ على سبيل المثال: [1.0 ،0.0 ،0.0]
- ميزة النقطة العائمة؛ على سبيل المثال: 8.3.

في هذه الحالة، سيتم تمثيل متجه الميزة لكل مثال بتسع قيم. نظرًا لقيم المثالفي القائمة السابقة، سيكون متجه الميزة:

```
0.0
1.0
0.0
0.0
0.0
0.0
0.0
```

```
1.0
8.3
```

Federated Learning

التعلم الموحد

نهج التعلم الآلي الموزع الذي يدرب (trains) نماذج (models) التعلم الآلي باستخدام الأمثلة اللامركزية الموجودة على الأجهزة مثل الهواتف الذكية.في التعلم الموحد، تقوم مجموعة فرعية من الأجهزة بتنزيل النموذج الحالي من خادم تنسيق مركزي. تستخدم الأجهزة الأمثلة المخزنة على الأجهزة لإجراء تحسينات على النموذج. تقوم الأجهزة بعد ذلك بتحميل تحسينات النموذج (ولكن ليس أمثلة التدريب) إلى الخادم المنسق، حيث يتم تجميعها مع تحديثات أخرى لإنتاج نموذج عالمي محسن. بعد التجميع، لم تعد هناك حاجة لتحديثات النموذج المحسوبة بواسطة الأجهزة، ويمكن تجاهلها.

نظرًا لأنه لا يتم تحميل أمثلة التدريب مطلقًا، فإن التعلم الموحد يتبع مبادئ الخصوصية الخاصة بجمع البيانات المركزة وتقليل البيانات.

Feedback loop

حلقة التغذية الخلفية

في التعلم الآلي، الموقف الذي تؤثر فيه تنبؤات النموذج على بيانات التدريب لنفس النموذج أو نموذج آخر. على سبيل المثال، سيؤثر النموذج الذي يوصي بالأفلام على الأفلام التي يشاهدها الأشخاص، والتي ستؤثر بعد ذلك على نماذج توصية الأفلام اللاحقة.

Feedforward neural network (FFN)

الشبكة العصبية امامية التغذية (FFN)

شبكة عصبية بدون اتصالات دورية أو متكررة. على سبيل المثال، الشبكات العصبية العميقة (deep neural networks) التقليدية هي شبكات عصبية امامية التغذية. على النقيض من الشبكات العصبية المتكررة (recurrent neural networks) التي تكون دورية.

Filter

الفلتر (المرشح)

هو التابع او المصفوفة التي تستخدمفي عملية الالتفاف (convolution operation) مع مصفوفة الادخال. يتم تمثيل الفلتر (المرشح) بواسطة متجه للأوزان نقوم بلف المدخلات. تعمل كل طبقة شبكة كمرشح لوجود ميزات أو أنماط محددة موجودةفي الصورة الأصلية.

لاكتشاف مرشح، من غير المناسب مكان وجود هذه الميزة أو النمط المحددفي الصورة الأصلية. تم تصميم هذه المرشحات خصيصًا لاكتشاف ما إذا كانت الصورة تحتوي على أي من هذه الخصائص أم لا. يتم إزاحة هذا المرشح عدة مرات ويتم تطبيقهفي مواضع مختلفة للصورة حتى تتم تغطية الصورة بأكملها بالتفصيل.

Few-shot learning

قليل من التعلم

نهج التعلم الآلي، الذي يستخدم غالبًا لتصنيف الكائنات، مصمم لتدريب المصنفات الفعالة من عدد قليل فقط من أمثلة التدريب.

انظر أيضًا التعلم بلقطة واحدة (one-shot learning).

Fine tuning

الضبط الدقيق

إجراء تحسين ثانوي لضبط معلمات نموذج (model) تم تدريبه بالفعل ليناسب مشكلة جديدة. غالبًا ما يشير الضبط الدقيق إلى إعادة ضبط أوزان نموذج مدرب غير خاضع للإشراف (unsupervised) إلى نموذج خاضع للإشراف (supervised).

Flatten layer

الطبقة المسطحة

تُستخدم الطبقة المسطحة (flatten year) لجعل الإدخال متعدد الأبعاد أحادي البعد، ويشيع استخدامهفي الانتقال من طبقة الالتفاف (convolution layer) إلى الطبقة المتصلة بالكامل (full connected layer).

Forget gate

بوابة النسيان

جزء من خلية الذاكرة طويلة قصيرة المدى (Long Short-Term Memory(LSTM)) التي تنظم تدفق المعلومات عبر الخلية. بوابة النسيان تحافظ على السياق من خلال تحديد المعلومات التي يجب تجاهلها من حالة الخلية.

Forward propagation

الانتشار الأمامي

الانتشار الأمامي هو طريقة الانتقال من طبقة الإدخال (يسارًا) إلى طبقة الإخراج (يمينًا)في الشبكة العصبية. تسمى عملية الانتقال من اليمين إلى اليسار، أي للخلف من المخرجات إلى طبقة الإدخال، بالانتشار الخلفي (backward propagation).

Full softmax

مرادف لـ softmax.

على النقيض من أخذ العينات المرشح (candidate sampling).

Fully connected layer

الطبقة المتصلة بالكامل

طبقة مخفية (hidden layer) تتصل فيها كل عقدة (node) بكل عقدةفي الطبقة المخفية اللاحقة.

تُعرف الطبقة المتصلة بالكامل أيضًا بطبقة كثيفة (dense layer).

F-Score

يجمع مقياس التقييم درجة F بين الدقة (precision) والاسترجاع (recall) كمقياس لفعالية التصنيف. يتم حسابه من حيث نسبة الأهمية الموزونة على الاسترجاع أو الدقة كما هو محدد بواسطة معامل β. ويسمى أيضا (F measure).

$$F\ measure = 2 \times (Recall \times Precision) / (\beta^2 \times Recall + Precision)$$

G

G

Gated Recurrent Unit (GRU)

وحدة البوابات المتكررة

تعد وحدة البوابات المتكررة (GRU) جزءًا من نموذج محدد للشبكة العصبية المتكررة (recurrent neural network) التي تهدف إلى استخدام الاتصالات من خلال سلسلة من العقد لأداء مهام التعلم الآلي المرتبطة بالذاكرة والتكتل. لها مزايا تفوق الذاكرة طويلة قصيرة المدى (LSTM). تستخدم GRU ذاكرة أقل وأسرع من LSTM، ومع ذلك، فإن LSTM أكثر دقة عند استخدام مجموعات البيانات ذات التسلسلات الأطول.

Generalization

التعميم

قدرة النموذج على عمل تنبؤات صحيحة على بيانات جديدة غير مرئية من قبل. النموذج الذي يمكن أن يعمم هو عكس النموذج الذي يعاني من الضبط الزائد(overfitting).

تقوم بتدريب نموذج على الأمثلة الموجودةفي مجموعة التدريب. وبالتالي، يتعلم النموذج خصائص البياناتفي مجموعة التدريب. يسأل التعميم بشكل أساسي عما إذاكان نموذجك يمكنه عمل تنبؤات جيدة بشأن الأمثلة غير الموجودةفي مجموعة التدريب.

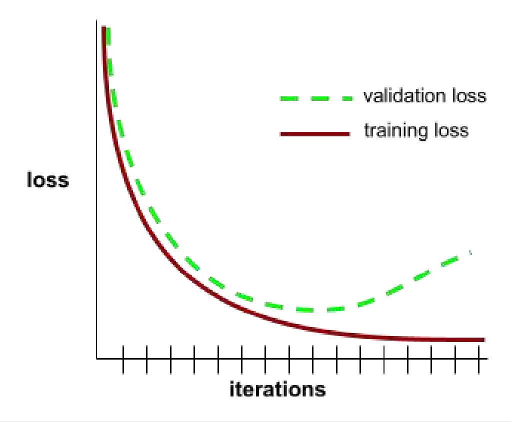
لتشجيع التعميم، يساعد التنظيم (regularization) النموذج على تدريب النموذج بشكل أقل دقة على خصوصيات البياناتفي مجموعة التدريب.

Generalization curve

منحنى التعميم

رسم بياني لكل من خطأ التدريب (training loss) وخطأ التحقق من الصحة (validation loss) كدالة لعدد التكرارات (iterations).

يمكن أن يساعدك منحنى التعميمفي اكتشاف الضبط الزائد (overfitting) المحتمل. على سبيل المثال، يشير منحنى التعميم التالي إلى الضبط الزائد لأن خطأ التحقق من الصحة يصبحفي النهاية أعلى بكثير من خطأ التدريب.



Generalized linear model

نموذج خطي معمم

تعميم لنماذج انحدار المربعات الصغرى (least squares regression)، والتي تعتمد على الضوضاء الغاوسية (Gaussian noise)، على أنواع أخرى من النماذج بناءً على أنواع أخرى من الضوضاء، مثل ضوضاء بواسون (Poisson noise) أو الضوضاء الفئوية. تتضمن أمثلة النماذج الخطية المعممة ما يلي:

- الانحدار اللوجستي (logistic regression).
- الانحدار متعدد الطبقات (multi-class regression).
- انحدار المربعات الأقل (least squares regression).

يمكن العثور على معلمات النموذج الخطي المعمم من خلال التحسين المحدب (convex optimization).

تعرض النماذج الخطية المعممة الخصائص التالية:

- متوسط التنبؤ لنموذج انحدار المربعات الصغرى الأمثل يساوي متوسط التسميةفي بيانات التدريب.
- متوسط الاحتمال الذي تنبأ به نموذج الانحدار اللوجستي الأمثل يساوي متوسط التسميةفي بيانات التدريب.

إن قوة النموذج الخطي المعمم محدودة بميزاته. على عكس النموذج العميق، لا يمكن للنموذج الخطي المعمم "تعلم ميزات جديدة".

Generative adversarial network (GAN)

شبكة الخصومة التوليدية

نظام لإنشاء بيانات جديدة يقوم فيها المولد (generator) بإنشاء بيانات ويحدد المميز (discriminator) ما إذا كانت البيانات التي تم إنشاؤها صالحة أم غير صالحة.

Generative model

النموذج التوليدي

من الناحية العملية، نموذج يقوم بأي مما يلي:

- ينشئ (يولد) أمثلة جديدة من مجموعة بيانات التدريب. على سبيل المثال، يمكن للنموذج التوليدي إنشاء الشعر بعد التدريب على مجموعة بيانات من القصائد. يندرج جزء المولد (generator) لشبكة الخصومة التوليدية (generative adversarial network(GAN)) في هذه الفئة.
- يحدد احتمال أن يأتي مثال جديد من مجموعة التدريب، أو أنه تم إنشاؤه من نفس الآلية التي أنشأت مجموعة التدريب. على سبيل المثال، بعد التدريب على مجموعة بيانات تتكون من جمل إنجليزية، يمكن للنموذج التوليدي تحديد احتمال أن يكون الإدخال الجديد جملة إنجليزية صالحة.

يمكن للنموذج التوليدي أن يميز نظريًا توزيع الأمثلة أو ميزات معينةفي مجموعة البيانات. هذا هو:

```
p(examples)
```

تعد نماذج التعلم غير الخاضعة للإشراف منتجة.

على النقيض من النماذج التمييزية (discriminative models).

Generative Pre-trained Transformer (GPT)

المحولات التوليدية المدربة مسبقًا

عائلة من نماذج اللغات الكبيرة المستندة إلى المحولات (Transformer) والتي طورتها شركة OpenAI.

يمكن أن تنطبق متغيرات GPT على طرائق متعددة، بمافي ذلك:

- توليد الصور (على سبيل المثال، ImageGPT)
- إنشاء نص إلى صورة (على سبيل المثال، DALL-E).

Generator

المولد

النظام الفرعي ضمن شبكة خصومة توليدية (generative adversarial network) تخلق أمثلة (examples) جديدة.

على النقيض من النموذج التمييزي (discriminative model).

Gini impurity

شائبة جيني

مقياس مشابه للإنتروبيا (entropy). تستخدم المقسمات (Splitters) قيمًا مشتقة من شوائب جيني أو إنتروبيا لتكوين شروط لأشجار قرار (decision trees) التصنيف. اكتساب المعلومات (Information gain) مشتق من الانتروبيا. لا يوجد مصطلح مكافئ مقبول عالميًا للمقياس المشتق من شوائب جيني؛ ومع ذلك، فإن هذا المقياس غير المسمى لا يقل أهمية عن اكتساب المعلومات.

تسمى شوائب جيني أيضًا بمؤشر جيني (gini index)، أو ببساطة جيني (gini).

Goodness of Fit

جودة اللائمة

يصف مدى ملاءمة نموذج ما مدى ملاءمته لمجموعة من المشاهدات (observations). عادةً ما تلخص مقاييس جودة الملاءمة التناقض بين القيم المرصودة (observed values) والقيم المتوقعة (expected values)في ظل النموذج.

GoogLeNet

GoogLeNet عبارة عن شبكة عصبية تلافيفية (convolutional neural network) يبلغ عمقها 22 طبقة. يمكنك تحميل نسخة محددة مسبقًا من الشبكة المدربة على مجموعات بيانات ImageNet أو Places365. تصنف الشبكة المدربة على ImageNet الصور إلى 1000 فئة كائن، مثل لوحة المفاتيح والماوس والقلم الرصاص والعديد من الحيوانات.

Gradient

التدرج

متجه المشتقات الجزئية (partial derivatives) بالنسبة لجميع المتغيرات المستقلة.في التعلم الآلي، التدرج الاشتقاقي هو متجه المشتقات الجزئية لدالة النموذج. يشير التدرج إلى اتجاه الصعود الأكثر انحدارًا.

Gradient boosting

تعزيز التدرج

خوارزمية تدريب حيث يتم تدريب النماذج الضعيفة لتحسين الجودة (تقليل الخطأ) بشكل متكرر لنموذج قوي. على سبيل المثال، يمكن أن يكون النموذج الضعيف نموذج شجرة قرار خطيًا أو صغيرًا. يصبح النموذج القوي مجموع كل النماذج الضعيفة التي تم تدريبها مسبقًا.

في أبسط أشكال تعزيز التدرج،في كل تكرار، يتم تدريب نموذج ضعيف للتنبؤ بتدرج الخطأ للنموذج القوي. بعد ذلك، يتم تحديث ناتج النموذج القوي بطرح التدرج المتوقع، على غرار التدرج الاشتقاقي (gradient descent).

تشمل الانواع الحديثة لتعزيز التدرج أيضًا المشتق الثاني (Hessian) للخسارةفي حساباتهم.

تُستخدم أشجار القرار (Decision trees) بشكل شائع كنماذج ضعيفةفي تعزيز التدرج.

Gradient boosted (decision) trees (GBT)

تعزيز التدرج لشجرة القرار

نوع من غابة القرار (decision forest) حيث:

- يعتمد التدريب (Training) على تعزيز التدرج (gradient boosting).
- النموذج الضعيف هو شجرة قرار (decision tree).

Gradient clipping

قص التدرج

آلية شائعة الاستخدام للتخفيف من مشكلة انفجار التدرج (exploding gradient problem) عن طريق تحديد (قص) القيمة القصوى للتدرجات بشكل مصطنع عند استخدام التدرج الاشتقاقي (gradient descent) لتدريب نموذج.

Gradient descent

التدرج الاشتقاقي

تقنية رياضية لتقليل الخطأ (loss). يعمل التدرج الاشتقاقي على تعديل الأوزان والتحيزات بشكل متكرر، وإيجاد أفضل مزيج تدريجيًا لتقليل الخطأ.

Graph

الرسم بياني

في (TensorFlow)، مواصفات حسابية. تمثل العقدفي الرسم البياني العمليات. يتم توجيه الحواف وتمثل تمرير نتيجة عملية (Tensor) كمعامل لعملية أخرى. استخدم (TensorBoard) لتصور رسم بياني.

Greedy policy

سياسة الجشع

في التعلم المعزز (reinforcement learning)، سياسة (policy) تختار دائمًا الإجراء الذي يحقق أعلى عائد (return) متوقع.

Ground truth

الحقيقة الأساسية

الواقع.

الشيء الذي حدث بالفعل.

على سبيل المثال، ضع في اعتبارك نموذج تصنيف ثنائي (binary classification) يتنبأ بما إذا كان الطالب في سنته الجامعية الأولى سيتخرج في غضون ست سنوات. الحقيقة الأساسية لهذا النموذج هي ما إذا كان هذا الطالب قد تخرج بالفعل في غضون ست سنوات أم لا.

H

H

Hadoop

Hadoop هو إطار معالجة مفتوح المصدر موزع يستخدم عندما يتعين علينا التعامل مع بيانات هائلة. يسمح لنا باستخدام قدرة المعالجة المتوازية للتعامل مع البيانات الضخمة (big data).

Hard Margin

الهامش الصلب

الهوامش الصلبةفي (SVMs) تعالج مشكلة التعميم للمدركين من خلال تعظيم الهامش، المحدد رسميًا على أنه الحد الأدنى للمسافة من حدود القرار (decision boundary) إلى نقاط التدريب.

Hallucination

هلوسة

إنتاج مخرجات تبدو معقولة ولكنها غير صحيحة من الناحية الواقعية من خلال نموذج توليدي (generative model) يزعم أنه يقوم بتأكيد حول العالم الحقيقي. على سبيل المثال، إذا ادعى وكيل (agent) الحوار أن باراك أوباما توفي عام 1865 ، فإن الوكيل يهلوس.

Hashing

تجزئة

في التعلم الآلي، توجد آلية لجمع البيانات الفئوية (categorical data)، لا سيما عندما يكون عدد الفئات كبيرًا، لكن عدد الفئات التي تظهر فعليًافي مجموعة البيانات صغير نسبيًا.

على سبيل المثال، الأرض هي موطن لحوالي 73000 نوع من الأشجار. يمكنك تمثيل كل نوع من أنواع الأشجار البالغ عددها 73000 نوعًافي 73000 مجموعة فئوية منفصلة. بدلاً من ذلك، إذا ظهر 200 نوع من هذه الأشجار فعليًافي مجموعة بيانات، فيمكنك استخدام التجزئة لتقسيم أنواع الأشجار إلى 500 مجموعة ربما.

يمكن أن تحتوي مجموعة واحدة على أنواع متعددة من الأشجار. على سبيل المثال، يمكن للتجزئة وضع الباوباب والقيقب الأحمر ـ وهما نوعان مختلفان وراثيًا ـ في نفس المجموعة. بغض النظر، لا تزال التجزئة طريقة جيدة لتعيين مجموعات فئوية كبيرةفي العدد المطلوب من المجموعات. يحول التجزئة ميزة فئوية لها عدد كبير من القيم الممكنة إلى عدد أقل بكثير من القيم عن طريق تجميع القيم بطريقة حتمية.

Heuristic

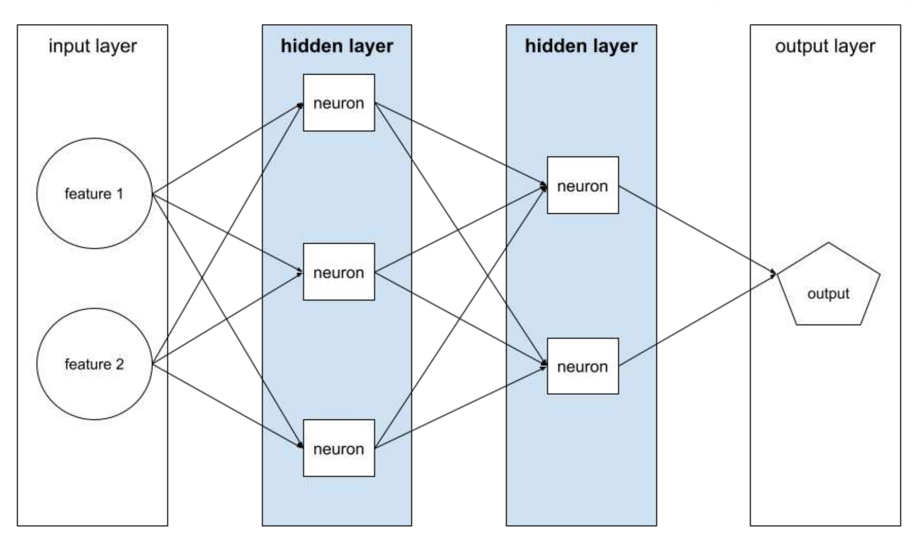
ارشادي

حل بسيط وسريع التنفيذ لمشكلة ما. على سبيل المثال، "من خلال الكشف عن مجريات الأمور، حققنا دقة تبلغ 86٪. عندما تحولنا إلى شبكة عصبية عميقة، ارتفعت الدقة إلى 98٪."

Hidden layer

الطبقة المخفية

طبقةفي الشبكة العصبية (neural network) بين طبقة الإدخال (input layer) (الميزات) وطبقة الإخراج (output layer) (التنبؤ). تتكون كل طبقة مخفية من خلية عصبية (neurons) واحدة أو أكثر. على سبيل المثال، تحتوي الشبكة العصبية التالية على طبقتين مخفيتين، الأولى بها ثلاث خلايا عصبية والثانية بها خليتان:



تحتوي الشبكة العصبية العميقة (deep neural network) على أكثر من طبقة مخفية. على سبيل المثال، الرسم التوضيحي السابق عبارة عن شبكة عصبية عميقة لأن النموذج يحتوي على طبقتين مخفيتين.

Hierarchical clustering

المجموعات الهرمية

فئة من خوارزميات التكتل (clustering) التي تنشئ شجرة عناقيد. التجميع الهرمي مناسب تمامًا للبيانات الهرمية، مثل التصنيفات النباتية. هناك نوعان من خوارزميات المجموعات الهرمية:

- يقوم التجميع التجميعي (Agglomerative clustering) أولاً بتعيين كل مثال إلى المجموعة الخاصة به، ويقوم بدمج المجموعات الأقرب بشكل متكرر لإنشاء شجرة هرمية.
- يجمع التجميع الانقسامي (Divisive clustering) أولاً جميع الأمثلةفي مجموعة واحدة ثم يقسم المجموعة بشكل متكرر إلى شجرة هرمية.

على النقيض من التجميع القائم على النقطة الوسطى (centroid-based clustering).

Hinge loss

خطأ هينج

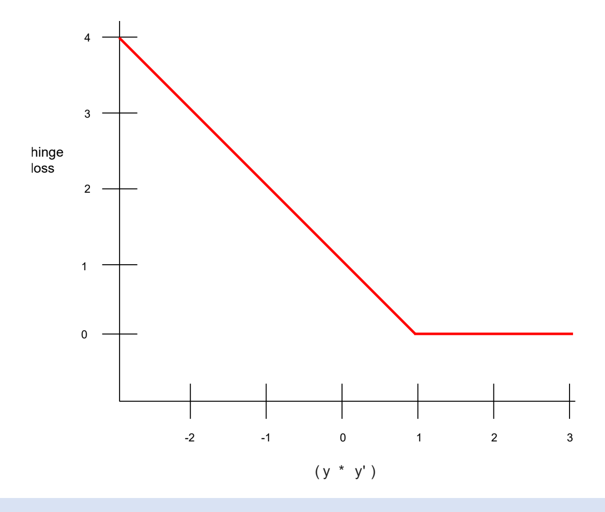
تم تصميم مجموعة دوال الخطأ (loss) للتصنيف (classification) لإيجاد حدود القرار (decision boundary) بعيدة قدر الإمكان عن كل مثال تدريبي، وبالتالي تعظيم الهامش بين الأمثلة والحدود. تستخدم KSVMs خطأ هينج (أو داله ذات صلة، مثل خطأ هينج التربيعية). بالنسبة للتصنيف الثنائي، يتم تعريف دالة خطأ هينج على النحو التالي:

$$\text{loss} = \max(0, 1 - (y * y'))$$

حيث y هي التسمية الحقيقية، إما 1- أو 1+ ، و y' هو الناتج الأولي لنموذج المصنف:

$$y' = b + w_1x_1 + w_2x_2 + \dots w_nx_n$$

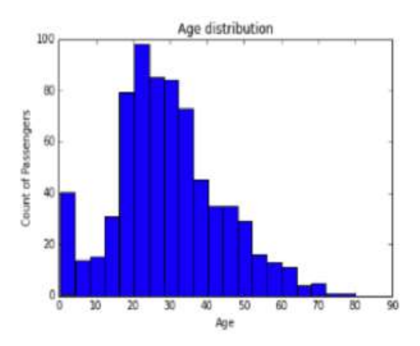
وبالتالي، فإن مخطط خطأ هينج مقابل (y * y') يبدو كما يلي:



Histogram

المدرج التكراري

المدرج التكراري هو أحد طرق رسم توزيع البيانات للمتغيرات المستمرة. على سبيل المثال، يوضح الشكل أدناه مدرجًا تكراريًا مع العمر على طول المحور السيني وتردد المتغير (عدد الركاب) على طول المحور الصادي.



Holdout data

بيانات الانتظار

أمثلة لم يتم استخدامها عن قصد (held out) أثناء التدريب. تعد مجموعة بيانات التحقق من الصحة (validation dataset) ومجموعة بيانات الاختبار (test dataset) أمثلة على بيانات الانتظار. تساعد بيانات الانتظارفي تقييم قدرة النموذج الخاص بك على التعميم على بيانات أخرى غير البيانات التي تم التدريب عليها. يوفر الخطافي مجموعة الانتظار تقديرًا أفضل للخطأ في مجموعة بيانات غير مرئية مقارنة بالخطأفي مجموعة التدريب.

Hyperparameter

المعلمات الفائقة

المتغيرات التي تقوم أنت أو خدمة ضبط المعلمة الفائقة بضبطها أثناء عمليات التدريب المتتالية للنموذج. على سبيل المثال، معدل التعلم (learning rate) هو معلمة فائقة. يمكنك ضبط معدل التعلم على 0.01 قبل جلسة تدريب واحدة. إذا حددت أن 0.01 مرتفع جدًا، فربما يمكنك ضبط معدل التعلم على 0.003 لجلسة التدريب التالية.

في المقابل (parameters)، المعلمات هي الأوزان (weights) المختلفة والتحيز (bias) الذي يتعلمه النموذج أثناء التدريب.

Hyperplane

المستوى الفائق

الحد الذي يفصل الفضاء إلى مساحتين فرعيين. على سبيل المثال، الخط هو مستوي فائقفي بعدين والمستوى هو مستوي فائقفي ثلاثة أبعاد. بشكل أكثر شيوعًافي التعلم الآلي، المستوى

الفائق هو الحد الفاصل بين الفضاء عالي الأبعاد. يستخدم كيرنل آلات المتجهات الداعمة (Kernel Support Vector Machines) المستويات الفائقة لفصل الفئات الإيجابية عن الفئات السلبية، غالبًافي مساحة عالية الأبعاد.

I

I

Image recognition

التعرف على الصور

عملية تصنف كائنًا (كائنات)، أو نمطًا (أنماطًا)، أو مفهومًا (مفاهيم)في صورة. يُعرف التعرف على الصور أيضًا باسم تصنيف الصور (image classification).

Imbalanced dataset

مجموعة بيانات غير متوازنة

مرادف لمجموعة البيانات غير المتوازنةفي الفئة (class-imbalanced dataset).

Implicit bias

التحيز الضمني

تكوين ارتباط أو افتراض تلقائيًا بناءً على النماذج الذهنية والذكريات. يمكن أن يؤثر التحيز الضمني على ما يلي:

- كيف يتم جمع البيانات وتصنيفها.
- كيف يتم تصميم أنظمة التعلم الآلي وتطويرها.

على سبيل المثال، عند إنشاء مصنف للتعرف على صور الزفاف، قد يستخدم المهندس وجود فستان أبيض في الصورة كميزة. ومع ذلك، كانت الفساتين البيضاء مألوفة فقط خلال عصور معينة وفي ثقافات معينة.

انظر أيضا التحيز التأكيدي (confirmation bias).

Imputation

التضمين

التضمين هو تقنية تستخدم لمعالجة القيم المفقودةفي البيانات. يتم ذلك إما عن طريق المقاييس الإحصائية مثل التضمين المتوسط / المنوال أو عن طريق تقنيات التعلم الآلي مثل احتساب خوارزمية أقرب الجيران (KNN).

فمثلا، إذا كانت البيانات على النحو التالي:

Name	Age
Akshay	23
Akshat	NA
Viraj	40

يحتوي الصف الثاني على قيمة مفقودة (NA)، لذلك نستخدم متوسط جميع الأعمار لتضمينها.

Name	Age
Akshay	23
Akshat	31.5
Viraj	40

Inception Network

شبكة الاستهلال

شبكة الاستهلال هي شبكة عصبية عميقة (deep neural network) ذات تصميم معماري يتكون من تكرار المكونات المشار إليها باسم وحدات الاستهلال (Inception modules).

Independently and identically distributed (i.i.d)

موزعة بشكل مستقل ومتماثل (i.i.d)

البيانات المستمدة من توزيع لا يتغير، وحيث لا تعتمد كل قيمة مرسومة على القيم التي تم رسمها سابقًا. i.i.d.هو الغاز المثالي للتعلم الآلي ـ بناء رياضي مفيد ولكنه تقريبًا لم يتم العثور عليه تمامًافي العالم الحقيقي. على سبيل المثال، قد يكون توزيع الزوار على صفحة الويب هو i.i.d. خلال فترة وجيزة من الوقت؛ أي أن التوزيع لا يتغير خلال تلك الفترة القصيرة وتكون زيارة شخص ما مستقلة بشكل عام عن زيارة شخص آخر. ومع ذلك، إذا قمت بتوسيع تلك الفترة الزمنية، فقد تظهر الاختلافات الموسميةفي زوار صفحة الويب.

انظر أيضا (nonstationarity).

Inference

الاستدلال

في التعلم الآلي، عملية صنع التنبؤات من خلال تطبيق نموذج مدرب على أمثلة غير مسماة (unlabeled examples).

الاستدلال له معنى مختلف إلى حد ما في الإحصاء.

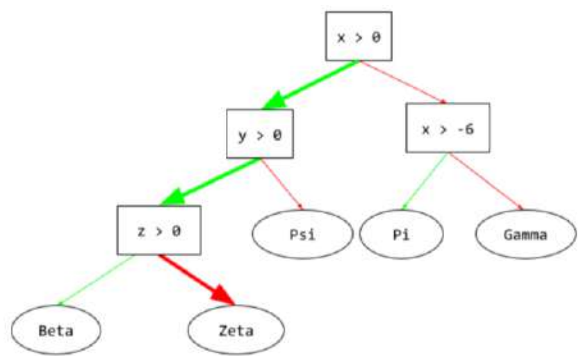
Inference path

مسار الاستدلال

في شجرة القرار (decision tree)، أثناء الاستدلال (inference)، المسار الذي يتخذه مثال معين من الجذر (root) إلى الشروط (conditions) الأخرى، وينتهي بورقة (leaf). على سبيل المثال،في شجرة القرار التالية، تُظهر الأسهم السميكة مسار الاستدلال لمثال بقيم الميزة التالية:

- x = 7
- y = 12
- z = -3

ينتقل مسار الاستدلالفي الرسم التوضيحي التالي عبر ثلاثة شروط قبل الوصول إلى الورقة (Zeta).



تظهر الأسهم الثلاثة السميكة مسار الاستدلال.

Information gain

اكتساب المعلومات

في غابات القرار (decision forests)، الفرق بين إنتروبيا العقدة والمجموع الموزون (بعدد الأمثلة) لانتروبيا العقد التابعة لها. إنتروبيا العقدة هي إنتروبيا الأمثلةفي تلك العقدة.

على سبيل المثال، ضعفي اعتبارك قيم الانتروبيا التالية:

- إنتروبيا العقدة الأم = 0.6.
- إنتروبيا عقدة فرعية واحدة مع 16 مثالاً ذا صلة = 0.2.

- إنتروبيا عقدة فرعية أخرى مع 24 مثالاً ذا صلة = 0.1.

إذن 40٪ من الأمثلة موجودةفي عقدة فرعية واحدة و60٪في العقدة الفرعية الأخرى. وبالتالي:

- مجموع الانتروبيا الموزون للعقد الفرعية = (0.4 * 0.2) + (0.6 * 0.1) = 0.14

إذن، اكتساب المعلومات هو:

- اكتساب المعلومات = إنتروبيا العقدة الأصل ـ مجموع الانتروبيا المرجحة للعقد الفرعية.
- اكتساب المعلومات = 0.6 ـ 0.14 = 0.46.

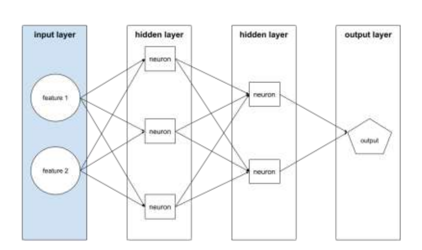
يسعى معظم المقسمين (splitters) إلى خلق ظروف تزيد من اكتساب المعلومات إلى أقصى حد.

Input layer

طبقة الإدخال

طبقة (layer) الشبكة العصبية (neural network) التي تحمل متجه الميزة (feature vector). أي أن طبقة الإدخال توفر أمثلة للتدريب (training) أو الاستدلال (inference).

على سبيل المثال، تتكون طبقة الإدخالفي الشبكة العصبية التالية من ميزتين:



Instance

مثال

نقطة بيانات أو صف أو عينةفي مجموعة بيانات.

مرادف لمثال (example) ولـ(observation).

Interpretability

قابلية التفسير

القدرة على شرح أو تقديم منطق نموذج تعلم الي بعبارات مفهومة للإنسان.

معظم نماذج الانحدار الخطي، على سبيل المثال، قابلة للتفسير بدرجة كبيرة. (ما عليك سوى إلقاء نظرة على الأوزان المدربة لكل ميزة.) كما يمكن تفسير غابات القرار بشكل كبير. ومع ذلك، تتطلب بعض النماذج تصورًا متطورًا لتصبح قابلة للتفسير.

Intersection over union (IoU)

تقاطع الاتحاد (IoU)

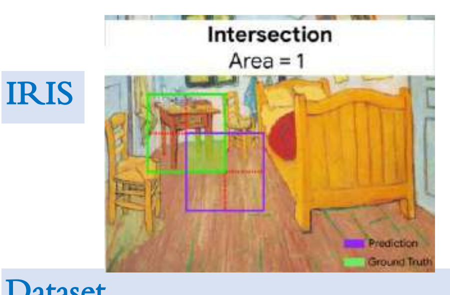
تقاطع مجموعتين مقسومة على اتحادهم.في مهام اكتشاف الصور بالتعلم الآلي، يتم استخدام IoU لقياس دقة مربع الإحاطة (bounding box) المتوقع للنموذج فيما يتعلق بمربع إحاطة الحقيقة الأساسية (ground-truth).في هذه الحالة، IoU للمربعين هي النسبة بين المساحة المتداخلة والمساحة الإجمالية، وتتراوح قيمتها من 0 (لا يوجد تداخل بين المربع المحيط المتوقع ومربع إحاطة الحقيقة الاساسية) إلى 1 (المربع المحيط المتوقع والاساسي ـ صندوق إحاطة الحقيقة له نفس الإحداثيات بالضبط).

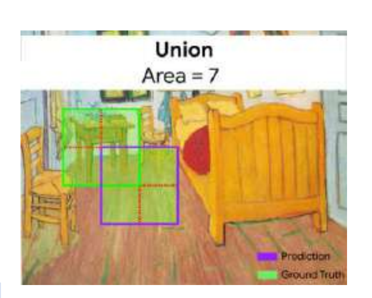
على سبيل المثالفي الصورة أدناه:

- تم تحديد مربع الإحاطة المتوقع (إحداثيات تحديد حيث يتنبأ النموذج بوجود الجدول الليليفي اللوحة) باللون الأرجواني.
- تم تحديد المربع المحيط بالحقيقة بالاساسية (الإحداثيات التي تحدد مكان وجود الطاولة الليليةفي اللوحة) باللون الأخضر.



هنا، تقاطع مربعات الاحاطة للتنبؤ والحقيقة الأساسية (أسفل اليسار) هو 1 ، واتحاد مربعات الاحاطة للتنبؤ والحقيقة الأساسية (أسفل اليمين) هو 7 ، لذا فإن IoU هي:





IRIS Dataset

مجموعة بيانات IRIS

IRIS (قاعدة بيانات لتصنيف شكل ثلاثة أنواع من الزهور).

Item matrix

مصفوفة العناصر

في أنظمة التوصية (recommendation systems)، مصفوفة من متجهات التضمين (embedding vectors) الناتجة عن عامل المصفوفة (Matrix factorization) الذي يحمل إشارات كامنة حول كل عنصر. يحتوي كل صف في مصفوفة العنصر على قيمة ميزة كامنة واحدة لجميع العناصر. على سبيل المثال، ضع في اعتبارك نظام توصية للأفلام. يمثل كل عمود في مصفوفة العنصر فيلمًا واحدًا. قد تمثل الإشارات الكامنة الأنواع، أو قد يكون من الصعب تفسير الإشارات التي تنطوي على تفاعلات معقدة بين النوع أو النجوم أو عمر الفيلم أو عوامل أخرى.

تحتوي مصفوفة العنصر على نفس عدد الأعمدة مثل المصفوفة المستهدفة التي يتم تحليلها. على سبيل المثال، نظرًا لنظام توصية الفيلم الذي يقيم 10000 عنوان فيلم، فإن مصفوفة العناصر ستحتوي على 10000 عمود.

Items

العناصر

في نظام التوصيات (recommendation systems)، الكيانات التي يوصي بها النظام. على سبيل المثال، مقاطع الفيديو هي العناصر التي يوصي بها متجر الفيديو، بينما الكتب هي العناصر التي يوصي بها متجر الكتب.

iteration

التكرار

تحديث واحد لمعلمات النموذج — أوزان (weights) النموذج وتحيزاته (biases) ـ أثناء التدريب. يحدد حجم الدُفعة (batch size) عدد الأمثلة التي يعالجها النموذجفي تكرار واحد. على سبيل المثال، إذاكان حجم الدُفعة هو 20، فإن النموذج يعالج 20 مثالًا قبل ضبط المعلمات.

عند تدريب شبكة عصبية (neural network)، يتضمن التكرار الفردي التمريرين التاليين:

- تمريرة أمامية لتقييم الخطأ على دفعة واحدة.
- الانتشار الخلفي(backpropagation) لضبط معلمات النموذج بناءً على الخطأ ومعدل التعلم.

J

J

Jupyter Notebook

Jupyter Notebook هو تطبيق الويب الأصلي لإنشاء المستندات الحسابية ومشاركتها. إنه يوفر تجربة بسيطة ومبسطة تتمحور حول المستندات.

K

K

Keras

كيراس

واجهة برمجة تطبيقات شائعة لتعلم الآلة بلغة Python. تعمل Keras على العديد من أطر التعلم العميق، بمافي ذلك TensorFlow، حيث يتم توفيرها كـ tf.keras.

Kernel

كيرنل (النواة)

هي المصفوفة التي تستخدمفي عملية الالتفاف (convolution operation)

Kernel Support Vector Machines (KSVMs)

كيرنل الة المتجهات الداعمة (KSVMs)

خوارزمية تصنيف تسعى إلى تعظيم الهامش بين الفئات الموجبة (posetive classes) والفئات السالبة (negative classes) عن طريق تعيين متجهات بيانات الإدخال إلى مساحة بُعد أعلى. على سبيل المثال، ضعفي اعتبارك مشكلة تصنيف تحتوي فيها مجموعة بيانات الإدخال على مئات الميزات. لتعظيم الهامش بين الفئات الإيجابية والسلبية، يمكن لـ KSVM تعيين هذه الميزات داخليًافي مساحة مليون بُعد. تستخدم KSVMs دالة خسارة تسمى خطأ هينج (hinge loss).

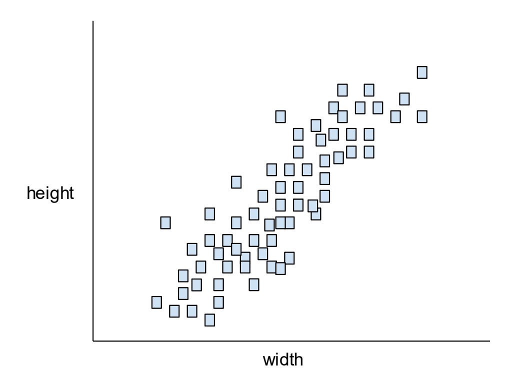
k-means

خوارزمية تجميع (clustering) شائعة تجمع الأمثلةفي التعلم غير الخاضع للإشراف. تقوم خوارزمية (k-mean) بشكل أساسي بما يلي:

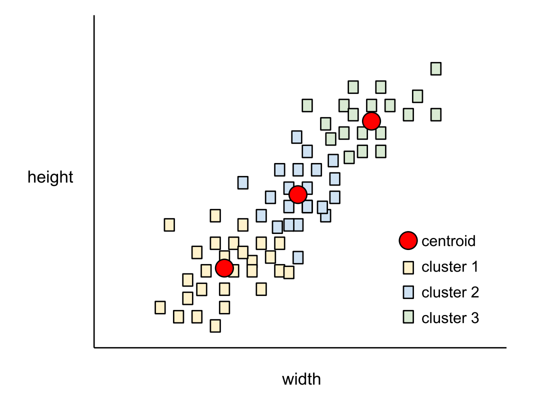
- يحدد بشكل تكراري أفضل k نقاط مركز (المعروفة باسم النقط الوسطى (centroids)).
- يعين كل مثال لأقرب النقطة الوسطى. هذه الأمثلة الأقرب إلى النقطة الوسطى نفسها تنتمي إلى نفس المجموعة.

تختار خوارزمية k-mean مواقع النقطة الوسطى لتقليل المربع التراكمي للمسافات من كل مثال إلى أقرب النقطة الوسطى.

على سبيل المثال، ضعفي اعتبارك المخطط التالية من الارتفاع إلى العرض:



إذا كانت k = 3، فإن خوارزمية k-mean ستحدد ثلاثة نقاط وسطى. يتم تعيين كل مثال لأقرب النقطة الوسطى، مما ينتج عنه ثلاث مجموعات:



تخيل أن الشركة المصنعة تريد تحديد الأحجام المثالية للسترات الصغيرة والمتوسطة والكبيرة للكلاب. تحدد النقط الوسطى الثلاثة متوسط الطول والعرض المتوسط لكل كلب في تلك المجموعة. لذلك، ربما ينبغي على الشركة المصنعة أن تبني أحجام السترات على تلك النقط الثلاثية. لاحظ أن النقطة الوسطى للكتلة عادة ليست مثالاً في الكتلة.

توضح الرسوم التوضيحية السابقة k-mean للأمثلة مع ميزتين فقط (الطول والعرض). لاحظ أن k-mean يمكنها تجميع الأمثلة عبر العديد من الميزات.

k-median

خوارزمية التجميع ترتبط ارتباطًا وثيقًا بـ (k-means). الفرق العملي بين الاثنين هو كما يلي:

- في k-mean، يتم تحديد النقطة الوسطى عن طريق تقليل مجموع مربعات المسافة بين مرشح النقطة الوسطى وكل من الأمثلة الخاصة به.

- في k-median، يتم تحديد النقط الوسطى عن طريق تقليل مجموع المسافة بين مرشح النقطة الوسطى وكل من الأمثلة الخاصة به.

k-nearest neighbors(KNN)

خوارزمية الجار الأقرب (KNN)

هو مصنف تعليمي غير معلمي خاضع للإشراف (supervised)، يستخدم التقارب (convergence) لعمل تصنيفات أو تنبؤات حول تجميع نقطة بيانات فردية. تفترض خوارزمية K-NN التشابه بين الحالة / البيانات الجديدة والحالات المتاحة وتضع الحالة الجديدةفي الفئة الأكثر تشابهًا مع الفئات المتاحة.

L

L

L_0 regularization

تنظيم L_0

نوع من التنظيم (regularization) يعاقب على العدد الإجمالي للأوزان (weights) غير الصفريةفي النموذج. على سبيل المثال، النموذج الذي يحتوي على 11 أوزانًا غير صفرية سيتم معاقبة أكثر من نموذج مشابه له 10 أوزان غير صفرية.

نادرا ما يتم استخدام تنظيم L_0.

يتم استخدام تنظيم L_1 ‏(L1 regularization) وتنظيم L_2 ‏(L2 regularization) بشكل أكبر بكثير من تنظيم L_0. ذلك لأن تنظيم L_1 و L_2 هي دوال محدبة (convex functions)، لكن تنظيم L_0 ليس دالة محدبة.

L_1 loss

خطأ L_1

دالة خطأ (loss function) تحسب القيمة المطلقة للفرق بين قيم التسمية (label) الفعلية والقيم التي يتوقعها النموذج (model). على سبيل المثال، إليك حساب خطأ L_1 لدفعة (batch) من خمسة أمثلة:

Actual value of example	Model's predicted value	Absolute value of delta
7	6	1
5	4	1
8	11	3
4	6	2
9	8	1
		8 = L_1 loss

تعتبر خطأ L_1 أقل حساسية للقيم المتطرفة (outliers) من خطأ L_2 ‏(L2 loss).

متوسط الخطأ المطلق (Mean Absolute Error) هو متوسط الخطأ L_1 لكل مثال.

L_1 regularization

تنظيم L_1

نوع من التنظيم (regularization) يعاقب على الأوزان بما يتناسب مع مجموع القيمة المطلقة للأوزان. يساعد تنظيم L_1في دفع أوزان الميزات غير الملائمة أو ذات الصلة بالكاد إلى الصفر بالضبط. تتم إزالة الميزة (feature) التي يبلغ وزنها 0 بشكل فعال من النموذج.

على النقيض من تنظيم L_2 (L_2 regularization).

L_2 loss

خطأ L_2

دالة خطأ (loss function) تحسب مربع الاختلاف بين قيم التسمية الفعلية والقيم التي يتوقعها النموذج (model). على سبيل المثال، إليك حساب خطأ L_2 لمجموعة من خمسة أمثلة:

Actual value of example	Model's predicted value	Square of delta
7	6	1
5	4	1
8	11	9
4	6	4
9	8	1
		16 = L_2 loss

بسبب التربيع، تضخم خطأ L_2 تأثير القيم المتطرفة (outliers). أي أن خطأ L_2 تتفاعل بشكل أقوى مع التنبؤات السيئة من خطأ L_1. على سبيل المثال، ستكون خطأ L_1 للدفعة السابقة 8 بدلاً من 16. لاحظ أن قيمة خارجية واحدة تمثل 9 من أصل 16.

تستخدم نماذج الانحدار (Regression models) عادةً خسارة L_2 كدالة خطأ.

متوسط الخطأ التربيعي (Mean Squared Error) هو متوسط خطأ L_2 لكل مثال. الخطأ التربيعي (Squared loss) هي اسم آخر لخطأ L_2.

L_2 regularization

تنظيم L_2

نوع من التنظيم (regularization) يعاقب على الأوزان (weights) بما يتناسب مع مجموع مربعات الأوزان. يساعد تنظيم L_2 على دفع الأوزان المتطرفة (outlier) (تلك ذات القيم الإيجابية العالية أو المنخفضة) الأقرب إلى 0 ولكن ليس تمامًا إلى 0. تظل الميزات ذات القيم القريبة جدًا من 0في النموذج ولكنها لا تؤثر على تنبؤ النموذج كثيرًا.

يعمل تنظيم L_2 دائمًا على تحسين التعميم في النماذج الخطية (linear models).

على النقيض من تنظيم L_1 (L_1 regularization).

Label

التسمية

في التعلم الآلي الخاضع للإشراف (supervised machine learning)، جزء "الإجابة" أو "النتيجة" من أحد الأمثلة (example).

يتكون كل مثال مسماة من واحد أو أكثر من الميزات (features) والتسمية (label). على سبيل المثال،في مجموعة بيانات اكتشاف الرسائل غير المرغوب فيها، من المحتمل أن يكون التصنيف إما "بريد عشوائي" أو "ليس بريداً عشوائيًا".في مجموعة بيانات هطول الأمطار، قد تكون التسمية هو كمية المطر التي سقطت خلال فترة معينة.

LaMDA (Language Model for Dialogue Applications)

نموذج اللغة لتطبيقات الحوار (LaMDA)

نموذج لغة كبير قائم (large language model) على المحولات (Transformer) طورته Google وتم تدريبه على مجموعة بيانات حوار كبيرة يمكنها توليد استجابات محادثة واقعية.

lambda

مرادف لمعدل التنظيم (regularization rate).

Lambda هو مصطلح زائد. نحن هنا نركز على تعريف المصطلح ضمن التنظيم (regularization).

landmarks

معالم

مرادف للنقاط المفتاحية (keypoints).

language model

نموذج لغة

نموذج (model) يقدر احتمالية وجود رمز أو تسلسل من الرموز المميزة (token)في تسلسل أطول من الرموز المميزة.

Large language model

نموذج لغة كبير

مصطلح غير رسمي بدون تعريف صارم يعني عادةً نموذج لغة (language model) يحتوي على عدد كبير من المعلمات. تحتوي بعض نماذج اللغات الكبيرة على أكثر من 100 مليار معلمة.

تعتمد معظم نماذج اللغات الكبيرة الحالية (على سبيل المثال، GPT) على بنية المحولات (Transformer).

Lasso Regression

انحدار لاسو

انحدار لاسو هي طريقة لتحليل الانحدار تقوم بالاختيار المتغير والتنظيم (regularization) من أجل تعزيز دقة التنبؤ وإمكانية تفسير النموذج الإحصائي الناتج.

يُطلق على نموذج الانحدار الذي يستخدم تقنية تنظيم L_1 (L_1 regularization) اسم انحدار لاسو والنموذج الذي يستخدم L_2 يسمى انحدار ريدج (ridge regression). الفرق الرئيسي بين هذين هو مصطلح العقوبة.

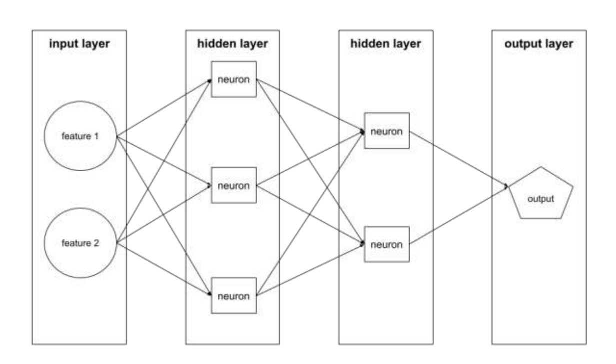
Layer

الطبقة

مجموعة من الخلايا العصبية (neurons)في شبكة عصبية (neural network). ثلاثة أنواع شائعة من الطبقات هي كما يلي:

- طبقة الإدخال (input layer)، والتي توفر قيمًا لجميع الميزات (features).
- طبقة مخفية (hidden layers) واحدة أو أكثر، والتي تجد العلاقات غير الخطية بين الميزات والتسمية.
- طبقة المخرجات (output layer) التي توفر التنبؤ.

على سبيل المثال، يوضح الرسم التوضيحي التالي شبكة عصبية ذات طبقة إدخال واحدة وطبقتين مخفيتين وطبقة إخراج واحدة:



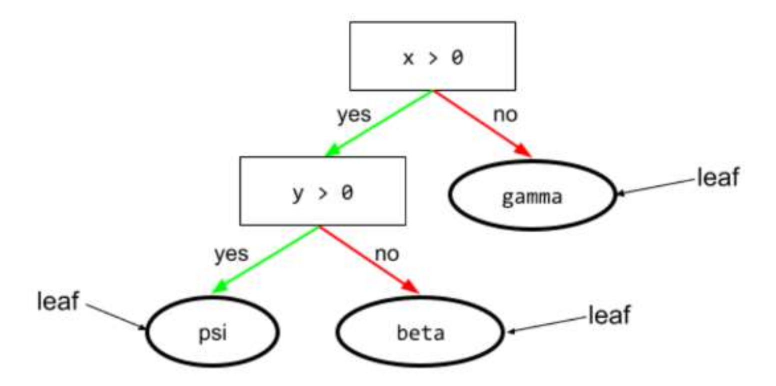
في TensorFlow، تعد الطبقات أيضًا دوال Python التي تأخذ خيارات الموترات (tensors) والتكوين كمدخلات وتنتج موترات أخرى كمخرجات.

Leaf

الورقة

أي نقطة نهايةفي شجرة القرار (decision tree). على عكس الشرط (condition)، لا تقوم الورقة بإجراء اختبار. بدلا من ذلك، ورقة هي توقع محتمل. الورقة هي أيضًا العقدة (node) الطرفية لمسار الاستدلال (inference path).

على سبيل المثال، تحتوي شجرة القرار التالية على ثلاث أوراق:



Learning rate

معدل التعلم

رقم النقطة العائمة الذي يخبر خوارزمية التدرج الاشتقاقي (gradient descent) مدى قوة تعديل الأوزان والتحيزاتفي كل تكرار (iteration). على سبيل المثال، معدل التعلم 0.3 يضبط الأوزان والتحيزات أقوى بثلاث مرات من معدل التعلم 0.1.

معدل التعلم هو معلم فائق (hyperparameter). إذا قمت بتعيين معدل التعلم منخفضًا جدًا، فسيستغرق التدريب وقتًا طويلاً. إذا قمت بتعيين معدل التعلم مرتفعًا جدًا، فغالبًا ما يواجه التدرج الاشتقاقي صعوبةفي الوصول إلى التقارب (convergence).

خلال كل تكرار، تقوم خوارزمية التدرج الاشتقاقي (gradient descent) بضرب معدل التعلم في التدرج (gradient). المنتج الناتج يسمى خطوة التدرج (gradient step).

Least squares regression

المربعات الدنيا للانحدار

نموذج انحدار خطي تم تدريبه عن طريق تقليل خطأ L_2 (L_2 Loss).

Lemmatization

الحصول على أصل الكلمة

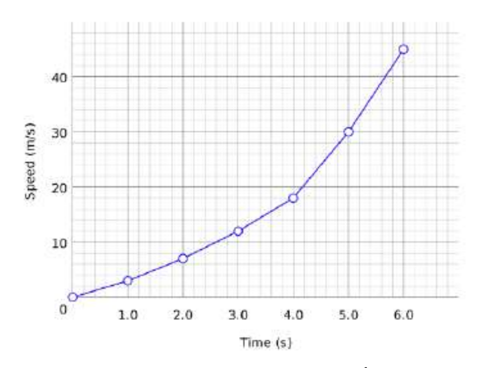
Lemmatizationفي اللغويات هي عملية تجميع الأشكال المنعكسة للكلمة معًا بحيث يمكن تحليلها كعنصر واحد ، محدد بواسطة lemma للكلمة ، أو نموذج القاموس.

Line Chart

المخطط الخطي

تُستخدم المخططات الخطية لعرض المعلومات كسلسلة من النقاط المتصلة بواسطة مقطع خط مستقيم. تُستخدم هذه المخططات لتوصيل المعلومات بصريًا، مثل إظهار زيادة أو نقصانفي الاتجاهفي البيانات عبر فترات زمنية.

في الرسم البياني أدناه، لكل حالة مرة، يتم عرض اتجاه السرعة والنقاط متصلة لعرض الاتجاه بمرور الوقت.



هذا المخطط لحالة واحدة. يمكن أيضًا استخدام المخططات الخطية لمقارنة التغييرات على مدار نفس الفترة الزمنية لحالات متعددة، مثل رسم سرعة الدراجة والسيارة والقطار بمرور الوقت في نفس المخطط.

Linear model

النموذج الخطي

نموذج (model) يعين وزنًا (weight) واحدًا لكل ميزة لعمل تنبؤات (predictions). (تتضمن النماذج الخطية أيضًا تحيزًا (bias)). وعلى النقيض من ذلك، فإن علاقة السمات بالتنبؤات في النماذج العميقة (deep models) تكون بشكل عام غير خطية (nonlinear).

عادة ما تكون النماذج الخطية أسهل في التدريب وأكثر قابلية للتفسير (interpretable) من النماذج العميقة. ومع ذلك، يمكن للنماذج العميقة تعلم العلاقات المعقدة بين الميزات.

الانحدار الخطي (Linear regression) والانحدار اللوجستي (logistic regression) نوعان من النماذج الخطية.

Linear

خطي

علاقة بين متغيرين أو أكثر يمكن تمثيلها فقط من خلال الجمع والضرب.

مخطط العلاقة الخطية عبارة عن خط.

على النقيض من غير الخطي (nonlinear).

Linear regression

الانحدار الخطي

نوع من نماذج التعلم الآلي يكون فيه كلا الأمرين التاليين صحيحًا:

- النموذج هو نموذج خطي (linear model).
- التوقع هو قيمة النقطة العائمة. (هذا هو جزء الانحدار (regression) من الانحدار الخطي.)

يتعارض الانحدار الخطي مع الانحدار اللوجستي (logistic regression). أيضا، مقارنة الانحدار (regression) مع التصنيف (classification).

Local Optima

الحل الأمثل المحلي

الحل الأمثل المحلي هو الحد الأقصى (الحد الأدنى أو الأقصى) لدالة الهدف (objective function) لمنطقة معينة من مساحة الإدخال، على سبيل المثال حوض في مشكلة التصغير (minimization).

Logistic regression

الانحدار اللوجيستي

نوع من نموذج الانحدار (regression model) يتنبأ باحتمالية. تتميز نماذج الانحدار اللوجستي بالخصائص التالية:

- التسمية فئوية (categorical). يشير مصطلح الانحدار اللوجستي عادةً إلى الانحدار اللوجستي الثنائي (binary logistic regression)، أي إلى نموذج يحسب الاحتمالات للتسميات ذات القيمتين المحتملتين. المتغير الأقل شيوعًا، الانحدار اللوجستي متعدد الحدود (multinomial logistic regression)، يحسب الاحتمالات للتسميات التي تحتوي على أكثر من قيمتين محتملتين.
- دالة الخطأ أثناء التدريب هي الخطأ اللوجيستي (Log Loss). (يمكن وضع وحدات الخطأ اللوجيستي المتعددة بالتوازي للتسميات التي تحتوي على أكثر من قيمتين محتملتين.)
- النموذج له بُنية خطية، وليس شبكة عصبية عميقة. ومع ذلك، فإن ما تبقى من هذا التعريف ينطبق أيضًا على النماذج العميقة (deep models) التي تتنبأ باحتمالات التسميات الفئوية.

على سبيل المثال، ضعفي اعتبارك نموذج الانحدار اللوجستي الذي يحسب احتمالية أن يكون البريد الإلكتروني المدخل بريدًا عشوائيًا أو ليس بريدًا عشوائيًا. أثناء الاستدلال، افترض أن النموذج يتوقع 0.72. لذلك، فإن النموذج يقدر:

- احتمال 72٪ أن يكون البريد الإلكتروني بريدًا عشوائيًا.
- فرصة بنسبة 28٪ ألا يكون البريد الإلكتروني بريدًا عشوائيًا.

يستخدم نموذج الانحدار اللوجستي البُنية التالية المكونة من خطوتين:

1. يولد النموذج تنبوًا أوليًا (y') من خلال تطبيق دالة خطية لميزات الإدخال.
2. يستخدم النموذج هذا التنبؤ الأولي كمدخل لدالة سيكمويد (sigmoid function)، والتي تحول التنبؤ الأولي إلى قيمة بين 0 و 1، حصريًا.

مثل أي نموذج انحدار، يتنبأ نموذج الانحدار اللوجستي بعدد. ومع ذلك، يصبح هذا الرقم عادةً جزءًا من نموذج التصنيف الثنائي على النحو التالي:

- إذا كان الرقم المتوقع أكبر من عتبة التصنيف (classification threshold)، فإن نموذج التصنيف الثنائي يتنبأ بالفئة الإيجابية.

- إذا كان الرقم المتوقع أقل من عتبة التصنيف، فإن نموذج التصنيف الثنائي يتنبأ بالفئة السلبية.

Logits

سجلات

متجه التنبؤات الأولية (غير المطبعة non-normalized) التي يولدها نموذج التصنيف، والتي يتم عادةً تمريرها بعد ذلك إلى دالة التطبيع (التسوية) (normalization). إذا كان النموذج يحل مشكلة تصنيف متعدد الفئات (multi-class classification)، فعادة ما تصبح السجلات مدخلاً لدالة softmax. تقوم دالة softmax بعد ذلك بإنشاء متجه للاحتمالات (المطبعة normalized) بقيمة واحدة لكل فئة ممكنة.

Log Loss

الخطأ اللوجيستي

دالة الخطأ المستخدمةفي الانحدار اللوجستي (logistic regression) الثنائي.

Long Short-Term Memory (LSTM)

الذاكرة طويلة قصيرة المدى (LSTM)

نوع من الخلايافي شبكة عصبية متكررة (recurrent neural network (RNN)) تُستخدم لمعالجة تسلسل البياناتفي تطبيقات مثل التعرف على خط اليد والترجمة الآلية والتعليق على الصور. تعالج LSTM مشكلة تلاشي التدرج (vanishing gradient problem) التي تحدث عند تدريب RNNs بسبب تسلسل البيانات الطويل من خلال الحفاظ على السجلفي حالة الذاكرة الداخلية بناءً على الإدخال والسياق الجديد من الخلايا السابقةفي RNN.

Loss

الخطأ

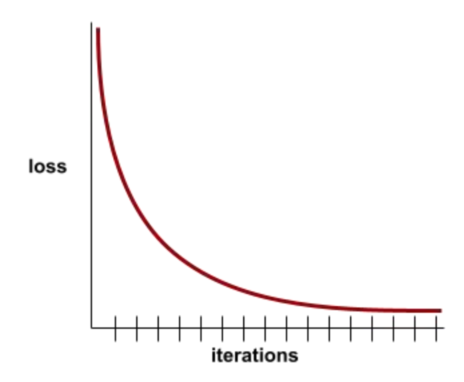
أثناء تدريب (training) نموذج خاضع للإشراف (supervised model)، مقياس لمدى بُعد تنبؤ (prediction) النموذج عن تسميته (label).

دالة الخطأ (loss function) تحسب الخطأ.

Loss curve

منحنى الخطأ

مخطط الخطأ (loss) كدالة لعدد التكرارات (iterations) التدريبية. يوضح الرسم البياني التالي منحنى خطأ نموذجي:



يمكن أن تساعدك منحنيات الخطأفي تحديد الوقت الذي يتقارب فيه نموذجك أو يتعدى حجمه.

يمكن لمنحنيات الخطأ رسم جميع أنواع الخطأ التالية:

- خطأ التدريب (training loss).
- خطأ التحقق من الصحة (validation loss).
- خطأ الاختبار (test loss).

انظر أيضا منحنى التعميم (generalization curve).

Loss function

دالة الخطأ

أثناء التدريب (training) أو الاختبار (testing)، دالة رياضية تحسب الخطأفي دفعة (batch) من الأمثلة. ترجع دالة الخطأ خطأ أقل للنماذج التي تقدم تنبؤات جيدة مقارنة بالنماذج التي تقوم بتنبؤات سيئة.

الهدف من التدريب عادةً هو تقليل الخطأ التي ترجعها دالة الخطأ.

توجد أنواع مختلفة من دوال الخطأ. اختر دالة الخطأ المناسبة لنوع النموذج الذي تقوم ببنائه. فمثلا:

- خطأ L_2 (L_2 loss) (أو متوسط الخطأ التربيعي (Mean Squared Error)) هي دالة الخطأ للانحدار الخطي (linear regression).

- الخطأ اللوجيستي (Log Loss) هو دالة الخطأ للانحدار اللوجستي (logistic regression).

Loss surface

سطح الخطأ

رسم بياني للوزن (weight) مقابل الخطأ (loss). يهدف التدرج الاشتقاقي (Gradient descent) إلى إيجاد الوزن (الأوزان) التي يكون سطح الخطأ فيها عند أدنى حد محلي.

M

M

Machine learning

تعلم الآلة

برنامج أو نظام يقوم بتدريب (trains) نموذج (model) من بيانات الإدخال. يمكن للنموذج المدرَّب أن يقدم تنبؤات مفيدة من بيانات جديدة (لم يسبق رؤيتها من قبل) مستمدة من نفس التوزيع مثل ذلك المستخدم لتدريب النموذج.

يشير التعلم الآلي أيضًا إلى مجال الدراسة المعني بهذه البرامج أو الأنظمة.

Machine Translation

الترجمة الالية

الترجمة الآلية، التي يشار إليها أحيانًا بالاختصار MT، ترجمة النص بواسطة خوارزمية مستقلة عن أي تدخل بشري.

Majority class

فئة الاغلبية

التسمية الأكثر شيوعًافي مجموعة البيانات غير المتوازنةفي الفئة (class-imbalanced dataset). على سبيل المثال، بالنظر إلى مجموعة بيانات تحتوي على 99٪ تسميات سلبية و1٪ تسميات إيجابية، فإن التسميات السلبية هي فئة الأغلبية.

على النقيض من فئة الأقلية (minority class).

Markov decision process (MDP)

قرارات عملية ماركوف

رسم بياني يمثل نموذج اتخاذ القرار حيث يتم اتخاذ القرارات (أو الإجراءات (actions)) للتنقل في سلسلة من الحالات (states) وفقًا للافتراض التي تحمله خاصية ماركوف (Markov property).في التعلم المعزز (reinforcement learning)، تعود هذه التحولات بين الحالات بمكافأة (reward) عددية.

Markov property

خاصية ماركوف

خاصية لبيئات (environments) معينة، حيث يتم تحديد انتقالات الحالة (state) بالكامل من خلال المعلومات المتضمنةفي الحالة الحالية وتصرف (action) الوكيل (agent).

Mask R-CNN

Mask R-CNN هي شبكة عصبية تلافيفية (CNN) وأحدثها من حيث تجزئة الصورة (image segmentation). يكتشف هذا البديل من الشبكة العصبية العميقة الكائناتفي صورة ما ويولد قناع تجزئة عالي الجودة لكل حالة.

Masked language model

نموذج اللغة المقنعة

نموذج لغوي (language model) يتنبأ باحتمالية أن تملأ الرموز المميزة (tokens) الفراغات في تسلسل. على سبيل المثال، يمكن لنموذج اللغة المقنعة حساب احتمالات الكلمة (الكلمات) المرشحة لتحل محل التسطيرفي الجملة التالية:

The ____ in the hat came back.

تستخدم الأدبيات عادةً السلسلة "MASK" بدلاً من التسطير. فمثلا:

The "MASK" in the hat came back.

معظم نماذج اللغة المقنعة الحديثة ثنائية الاتجاه (bidirectional).

matplotlib

مكتبة رسم python ثنائية الابعاد مفتوحة المصدر. يساعدك matplotlib على رسم الجوانب المختلفة للتعلم الآلي.

Matrix factorization

في الرياضيات، آلية لإيجاد المصفوفات التي يقترب حاصل ضربها النقطي (dot product) من المصفوفة المستهدفة.

Matrix Multiplication

الضرب المصفوفي

في الرياضيات، لا سيمافي الجبر الخطي (linear algebra)، يعد ضرب المصفوفة عملية ثنائية تنتج مصفوفة من مصفوفتين. لضرب المصفوفة، يجب أن يكون عدد الأعمدةفي المصفوفة الأولى مساويًا لعدد الصفوففي المصفوفة الثانية.

Max Pooling

التجميع حد أكثر

هي عملية تجميع (pooling) تحسب القيمة القصوى لتصحيحات خريطة الميزات (feature map)، وتستخدمها لإنشاء خريطة معالم مختزلة downsampled (مجمعة pooled). يستخدم عادة بعد الطبقة التلافيفية (Convolution layer).

Mean

المتوسط

بالنسبة لمجموعة البيانات (data set)، يُقال إن المتوسط هو متوسط قيمة جميع الأرقام. يمكن استخدامه أحيانًا كتمثيل للبيانات بأكملها.

على سبيل المثال، إذا كانت لديك علامات الطلاب من الفصل، وسألت عن مدى جودة أداء الفصل. لن يكون من المهم أن تقول علامات كل طالب على حدة، بدلاً من ذلك، يمكنك العثور على متوسط الفصل، والذي سيكون ممثلًا لأداء الفصل.

للعثور على المتوسط، اجمع كل الأرقام ثم اقسم على عدد العناصرفي المجموعة.

على سبيل المثال، إذا كانت الأرقام هي 1،2،3،4،5،6،7،8،8 فإن المتوسط سيكون 9/44 = 4.89.

Median

الوسيط

عادةً ما يكون الوسيط مجموعة من الأرقام هو القيمة المتوسطة. عندما تكون الأرقام الإجماليةفي المجموعة زوجية، سيكون الوسيط هو متوسط القيمتين الوسطيتين. الوسيط يستخدم لقياس الاتجاه المركزي.

لحساب الوسيط لمجموعة من الأرقام، اتبع الخطوات التالية:

1. رتب الأرقام بترتيب تصاعدي أو تنازلي
2. أوجد القيمة الوسطى، والتي ستكون n / 2 (حيث n هي الأرقامفي المجموعة).

Mean Absolute Error (MAE)

متوسط الخطأ المطلق

متوسط الخطأ لكل مثال عند استخدام خطأ L_1 (L_1 loss). احسب متوسط الخطأ المطلق كما يلي:

1. احسب خطأ L_1 للدفعة (batch).

2. اقسم خطأ L_1 على عدد الأمثلة في الدفعة.

على سبيل المثال، ضع في اعتبارك حساب خطأ L_1 على الدفعة التالية المكونة من خمسة أمثلة:

Actual value of example	Model's predicted value	Loss (difference between actual and predicted)
7	6	1
5	4	1
8	11	3
4	6	2
9	8	1
		8 = L_1 loss

إذن، خطأ L_1 هي 8 وعدد الأمثلة 5. لذلك، فإن متوسط الخطأ المطلق (MAE) هو:

```
Mean Absolute Error = L₁ loss / Number of Examples
Mean Absolute Error = 8/5 = 1.6
```

تتعارض متوسط للخطأ المطلق (MAE) مع الخطأ التربيعي المتوسط (Mean Squared Error) وجذر الخطأ التربيعي المتوسط (Root Mean Squared Error).

Mean Squared Error (MSE)

الخطأ التربيعي المتوسط

متوسط الخطأ لكل مثال عند استخدام خطأ L_2 (L_2 loss). احسب الخطأ التربيعي المتوسط على النحو التالي:

- احسب خطأ L_2 للدفعة.
- اقسم خطأ L_2 على عدد الأمثلة في الدفعة.

على سبيل المثال، ضع في اعتبارك الخطأ في الدفعة التالية المكونة من خمسة أمثلة:

Actual value	Model's prediction	Loss	Squared loss
7	6	1	1
5	4	1	1
8	11	3	9
4	6	2	4
9	8	1	1
			16 = L_2 loss

لذلك، فإن الخطأ التربيعي المتوسط هو:

```
Mean Squared Error = L₂ loss / Number of Examples
Mean Squared Error = 16/5 = 3.2
```

يعد الخطأ التربيعي المتوسط محسن (optimizer) تدريب شائع، خاصةً للانحدار الخطي (linear regression).

يتباين الخطأ التربيعي المتوسط مع متوسط الخطأ المطلق (Mean Absolute Error) وجذر الخطأ التربيعي المتوسط (Root Mean Squared Error).

يستخدم TensorFlow Playground الخطأ التربيعي المتوسط لحساب قيم الخطأ.

Metric

المقياس

هي إحصائية تهمك.

الهدف (objective) هو مقياس يحاول نظام التعلم الآلي تحسينه.

Mini-batch

دفعة صغيرة

مجموعة فرعية صغيرة مختارة عشوائيًا من الدُفعة (batch) تمت معالجتهافي تكرار (iteration) واحد. عادة ما يكون حجم الدُفعة (batch size) الخاصة بالدُفعة الصغيرة بين 10 و1000 مثال.

على سبيل المثال، افترض أن مجموعة التدريب بأكملها (الدفعة الكاملة) تتكون من 1000 مثال. افترض أيضًا أنك قمت بتعيين حجم الدُفعة لكل دفعة صغيرة على 20. لذلك، يحدد كل تكرار الخطأفي عشرين عشوائيًا من أصل 1000 مثال ثم يعدل الأوزان (weights) والتحيزات (biases) وفقًا لذلك.

يعتبر حساب الخطأ على دفعة صغيرة أكثر فاعلية من حساب الخطأفي جميع الأمثلةفي الدفعة الكاملة.

Mini-batch stochastic gradient descent

التدرج الاشتقاقي العشوائي مع الدفعات الصغيرة

خوارزمية تدرج اشتقاقي (gradient descent) تستخدم دفعات صغيرة (mini-batches). بمعنى آخر، يقدر التدرج الاشتقاقي العشوائي للدفعة الصغيرة التدرج بناءً على مجموعة فرعية صغيرة من بيانات التدريب. يستخدم التدرج الاشتقاقي العشوائي المنتظم (stochastic gradient descent) دفعة صغيرة من الحجم 1.

Minimax loss

خطأ minimax

دالة خطأ لشبكات الخصومة التوليدية (generative adversarial networks)، بناءً على الانتروبيا المتقاطعة (cross-entropy) بين توزيع البيانات المتولدة والبيانات الحقيقية.

Minority Class

فئة الأقلية

التسمية الأقل شيوعًافي مجموعة البيانات غير المتوازنةفي الفئة (class-imbalanced dataset). على سبيل المثال، بالنظر إلى مجموعة بيانات تحتوي على 99٪ تسميات سلبية و1٪ تسميات إيجابية، فإن التسميات الإيجابية هي فئة الأقلية.

على النقيض من فئة الأغلبية (majority class).

مجموعة التدريب التي تحتوي على مليون مثال تبدو مثيرة للإعجاب. ومع ذلك، إذا كان تمثيل فئة الأقلية ضعيفًا، فقد تكون مجموعة التدريب الكبيرة جدًا غير كافية. ركز بشكل أقل على العدد الإجمالي للأمثلةفي مجموعة البيانات وأكثر على عدد الأمثلةفي فئة الأقلية.

إذاكانت مجموعة البيانات الخاصة بك لا تحتوي على أمثلة كافية لفئة الأقليات، ففكرفي استخدام الاختزال (downsampling) لتكملة فئة الأقلية.

Missing value

القيم المفقودة

قيمة الميزة (feature) غير معروفة أو غير موجودة. هناك عدة أسباب محتملة لفقد القيمة، مثل: لم يتم قياسها؛ كان هناك عطل في الجهاز؛ السمة لا تنطبق، أو لا يمكن معرفة قيمة السمة. بعض الخوارزميات لديها مشاكل في التعامل مع القيم المفقودة.

ML

اختصار للتعلم الآلي (machine learning).

MNIST

مجموعة بيانات عامة جمعتها LeCun و Cortes و Burges تحتوي على 60.000 صورة، كل صورة توضح كيف كتب الإنسان يدويًا رقمًا معينًا من 0 إلى 9. يتم تخزين كل صورة كمصفوفة x2828 من الأعداد الصحيحة، حيث يكون كل عدد صحيح عبارة عن قيمة تدرج الرمادي بين 0 و 255، ضمناً.

MNIST عبارة عن مجموعة بيانات أساسية للتعلم الآلي ، وغالبًا ما تُستخدم لاختبار مناهج التعلم الآلي الجديدة.

MobileNet

موبايل نت

موبايل نت هو نوع من الشبكات العصبية التلافيفية (CNN) المصممة لتطبيقات الرؤية المتنقلة والمدمجة. وهي تستند إلى بنية مبسطة تستخدم تلافيفات قابلة للفصل في العمق لبناء شبكات عصبية عميقة خفيفة الوزن يمكن أن يكون لها زمن انتقال منخفض للأجهزة المحمولة والأجهزة المدمجة.

Mode

المنوال

المنوال هو القيمة الأكثر شيوعًا التي تحدث في المجتمع. إنه مقياس لقياس الاتجاه المركزي، مثلا طريقة للتعبير، في رقم فردي (عادةً)، عن معلومات مهمة حول متغير عشوائي أو مجموعة سكانية.

يمكن حساب الوضع باستخدام الخطوات التالية:

1. احسب عدد مرات ظهور كل قيمة.
2. خذ القيمة التي تظهر أكثر.

Modality

الأسلوب

فئة بيانات عالية المستوى. على سبيل المثال، الأرقام والنصوص والصور والفيديو والصوت خمس طرق مختلفة.

Model

النموذج

بشكل عام، أي بناء رياضي يعالج بيانات الإدخال ويعيد المخرجات. بصيغة مختلفة، النموذج هو مجموعة المعلمات والهيكل اللازمين للنظام لإجراء تنبؤات. في التعلم الآلي الخاضع للإشراف (supervised machine learning)، يأخذ النموذج مثالاً (example) كمدخل ويستنتج التنبؤ (prediction) كمخرج. ضمن التعلم الآلي الخاضع للإشراف، تختلف النماذج إلى حد ما. فمثلا:

- يتكون نموذج الانحدار الخطي من مجموعة من الأوزان (weights) والانحياز (bias).
- يتكون نموذج الشبكة العصبية (neural network) من:
  - مجموعة من الطبقات المخفية (hidden layers) تحتوي كل منها على خلية عصبية (neuron) واحدة أو أكثر.
  - الأوزان والتحيز المرتبط بكل خلية عصبية.
- يتكون نموذج شجرة القرار (decision tree) من:
  - شكل الشجرة أي النمط الذي ترتبط به الشروط والأوراق.
  - الشروط والأوراق.

يمكنك حفظ النموذج أو استعادته أو عمل نسخ منه.

ينشئ التعلم الآلي غير الخاضع للإشراف (Unsupervised machine learning) أيضًا نماذج، وعادةً ما تكون دالة يمكنها تعيين مثال إدخال إلى التكتل (cluster) الأكثر ملاءمة.

Model deployment

نشر النموذج

استخدام النموذج المتعلم. عادةً ما يشير نشر النموذج إلى تطبيق النموذج على بيانات حقيقية.

Model capacity

سعة النموذج

تعقيد المشكلات التي يمكن أن يتعلمها النموذج. كلما زادت تعقيد المشكلات التي يمكن أن يتعلمها النموذج، زادت سعة النموذج. عادة ما تزداد قدرة النموذج مع عدد معلمات النموذج. للحصول على تعريف رسمي لسعة المصنف،

Model parallelism

توازي النموذج

طريقة لتوسيع نطاق التدريب أو الاستدلال تضع أجزاء مختلفة من نموذج واحد على أجهزة مختلفة. يعمل توازي النموذج على تمكين النماذج الأكبر من أن تتناسب مع جهاز واحد.

راجع أيضًا توازي البيانات (data parallelism).

Model Selection

اختيار النموذج

اختيار النموذج هو مهمة اختيار نموذج إحصائي من مجموعة من النماذج المعروفة. الطرق المختلفة التي يمكن استخدامها لاختيار النموذج هي:

- تحليل البيانات استكشافية (EDA).
- الأساليب العلمية.

Model training

تدريب النموذج

عملية تحديد النموذج (model) الأفضل.

Model Validation

التحقق من صحة النموذج

يستلزم التحقق من صحة النموذج الخطوات التالية:

1. ملاءمة (Fitting) النموذج لمجموعة من بيانات التدريب.
2. ضبط المعلمات الفائقة (hyperparameters) للنموذج باستخدام مجموعة من بيانات التحقق من الصحة.
3. تقييم الأداء التنبئي (predictive performance) على مجموعة بيانات اختبار مستقلة.

الأسلوبان الأكثر شيوعًا للتحقق هما:

1. تقسيم البيانات إلى مجموعة تدريب وتحقق واختبار.
2. استخدام التحقق المتقاطع (cross-validation)، حيث يتم تدريب النموذج على مجموعات فرعية مختلفة من البيانات.

Momentum

الزخم

خوارزمية تدرج اشتقاقي (gradient descent) معقدة تعتمد فيها خطوة التعلم ليس فقط على المشتق في الخطوة الحالية، ولكن أيضًا على مشتقات الخطوة (الخطوات) التي سبقتها مباشرة. يتضمن الزخم حساب متوسط متحرك أسيًا للتدرجات بمرور الوقت، مشابهًا للزخم في الفيزياء. أحيانًا يمنع الزخم التعلم من الوقوع في الحدود الدنيا المحلية.

Multi-class classification

التصنيف متعدد الفئات

في التعلم الخاضع للإشراف، وهي مشكلة تصنيف (classification) تحتوي فيها مجموعة البيانات على أكثر من فئتين (classes) من التسميات (labels). على سبيل المثال، يجب أن تكون التسمياتفي مجموعة بيانات Iris واحدة من الفئات الثلاثة التالية:

- Iris setosa
- Iris virginica
- Iris versicolor

نموذج تم تدريبه على مجموعة بيانات Iris التي تتنبأ بنوع Iris على أمثلة جديدة يقوم بتصنيف متعدد الفئات.

في المقابل، فإن مشاكل التصنيف التي تميز بين فئتين بالضبط هي نماذج تصنيف ثنائية (binary classification models). على سبيل المثال، نموذج البريد الإلكتروني الذي يتنبأ بالبريد العشوائي أو ليس البريد العشوائي هو نموذج تصنيف ثنائي.

في مشاكل التجميع، يشير التصنيف متعدد الفئات إلى أكثر من مجموعتين.

Multi-class logistic regression

الانحدار اللوجستي متعدد الفئات

استخدام الانحدار اللوجستي (logistic regression)في مسائل التصنيف متعدد الفئات (multi-class classification).

Multi-head self-attention

الانتباه الذاتي متعدد الرؤوس

امتداد للانتباه الذاتي (self-attention) يطبق آلية الانتباه الذاتي عدة مرات لكل موضعفي تسلسل الإدخال.

قدمت المحولات (Transformers) الانتباه الذاتي متعدد الرؤوس.

Multimodal model

نموذج متعدد الوسائط

نموذج تشتمل مدخلاته و / أو مخرجاته على أكثر من أسلوب (modality). على سبيل المثال، ضعفي اعتبارك نموذجًا يأخذ كلاً من الصورة والتعليق النصي (طريقتان) كميزات (features)، ويخرج درجة تشير إلى مدى ملاءمة التسمية التوضيحية النصية للصورة. لذا، فإن مدخلات هذا النموذج متعددة الوسائط والمخرجات أحادية الوسائط.

Multinomial classification

مرادف لتصنيف متعدد الفئات (multi-class classification).

Multinomial regression

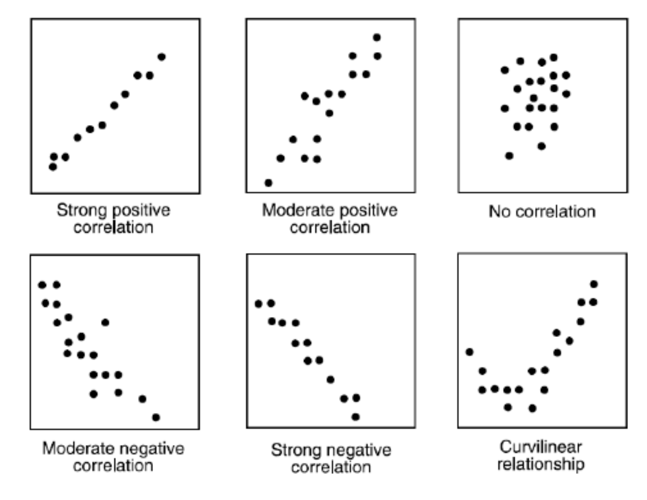
مرادف للانحدار اللوجستي متعدد الفئات (multi-class logistic regression).

Multivariate Analysis

التحليل متعدد المتغيرات

التحليل متعدد المتغيرات هو عملية مقارنة وتحليل تبعية المتغيرات المتعددة على بعضها البعض.

على سبيل المثال، يمكننا إجراء تحليل ثنائي المتغير لمزيج من ميزتين مستمرتين وإيجاد علاقة بينهما.



Multivariate Regression

الانحدار متعدد المتغيرات

يُطلق على نموذج الانحدار المصمم للتعامل مع المتغيرات التابعة المتعددة نموذج الانحدار متعدد المتغيرات.

MxNet

MXNet هو إطار عمل برمجي للتعلم العميق مفتوح المصدر ، يستخدم لتدريب ونشر الشبكات العصبية العميقة. إنه قابل للتطوير، مما يسمح بتدريب نموذج سريع ويدعم نموذج برمجة مرن ولغات برمجة متعددة.

N

N

Naive Bayes

نايف بايز

إنها تقنية تصنيف تعتمد على نظرية بايز (Bayes’ theorem) مع افتراض الاستقلال بين المتنبئين. بعبارات بسيطة، يفترض مصنف نايف بايز أن وجود ميزة معينةفي فئة لا علاقة لها بوجود أي ميزة أخرى. على سبيل المثال، يمكن اعتبار الفاكهة على أنها تفاحة إذا كانت حمراء ودائرية وقطرها حوالي 3 بوصات. حتى إذا كانت هذه الميزات تعتمد على بعضها البعض أو على وجود ميزات أخرى، فإن مصنف نايف بايز سوف يأخذفي الاعتبار كل هذه الخصائص للمساهمة بشكل مستقلفي احتمال أن تكون هذه الفاكهة تفاحة.

Named entity recognition (NER)

التعرف على الكيان المحدد

التعرف على الكيان المحدد هو مهمة فرعية لاستخراج المعلومات التي تسعى إلى تحديد وتصنيف الكيانات المسماة المذكورةفي نص غير منظم إلى فئات محددة مسبقًا مثل أسماء الأشخاص والمؤسسات والمواقع والأكواد الطبية والتعبيرات الزمنية والكميات والقيم النقدية والنسب المئوية وما إلى ذلك.

NaN

NaN تعني "ليس رقمًا". إنها قيمة نوع بيانات رقمية تمثل قيمة غير محددة أو غير قابلة للتمثيل. إذا كانت مجموعة البيانات تحتوي على قيم NaNفي مكان ما، فهذا يعني أن البيانات الموجودةفي هذا الموقع إما مفقودة أو ممثلة بشكل غير صحيح.

NaN trap

عندما يتحول رقم واحدفي نموذجك إلى NaN أثناء التدريب، مما يتسببفي تحويل العديد من الأرقام الأخرىفي نموذجك أو جميعها إلى NaNفي النهاية.

NaN هو اختصار لـ Not a Number.

Natural Language Processing (NLP)

معالجة اللغة الطبيعية

بكلمات بسيطة، معالجة اللغة الطبيعية هي مجال يهدف إلى جعل أنظمة الكمبيوتر تفهم كلام الإنسان. يتكون معالجة اللغة الطبيعية من تقنيات لمعالجة النص الخام وهيكله وتصنيفه واستخراج المعلومات.

ChatBot هو مثال كلاسيكي على معالجة اللغة الطبيعية، حيث تتم معالجة الجمل وتنظيفها وتحويلها إلى تنسيق مفهوم آليًا.

Natural Language Toolkit (NLTK)

مكتبة صندوق أدوات اللغات الطبيعية

مجموعة أدوات اللغة الطبيعية، أو بشكل أكثر شيوعًا NLTK، هي مجموعة من المكتبات والبرامج لمعالجة اللغة الطبيعية الرمزية والإحصائية للغة الإنجليزية المكتوبة بلغة برمجة Python.

Natural language understanding

فهم اللغة الطبيعية

تحديد نوايا المستخدم بناءً على ما كتبه أو قاله المستخدم. على سبيل المثال، يستخدم محرك البحث فهم اللغة الطبيعية لتحديد ما يبحث عنه المستخدم بناءً على ما كتبه المستخدم أو قاله.

Negative class

الفئة السلبية

في التصنيف الثنائي (binary classification)، يُطلق على إحدى الفئات اسم إيجابي ويطلق على الآخر اسم سلبي. الفئة الإيجابية هي الشيء أو الحدث الذي يختبر النموذج من أجله والفئة السلبية هي الاحتمال الآخر. فمثلا:

- قد تكون الفئة السلبيةفي الاختبار الطبي "ليست ورمًا".
- قد تكون الفئة السلبيةفي مصنف البريد الإلكتروني "ليست بريداً عشوائيًا".

على النقيض من الفئة الإيجابية (positive class).

Neural network

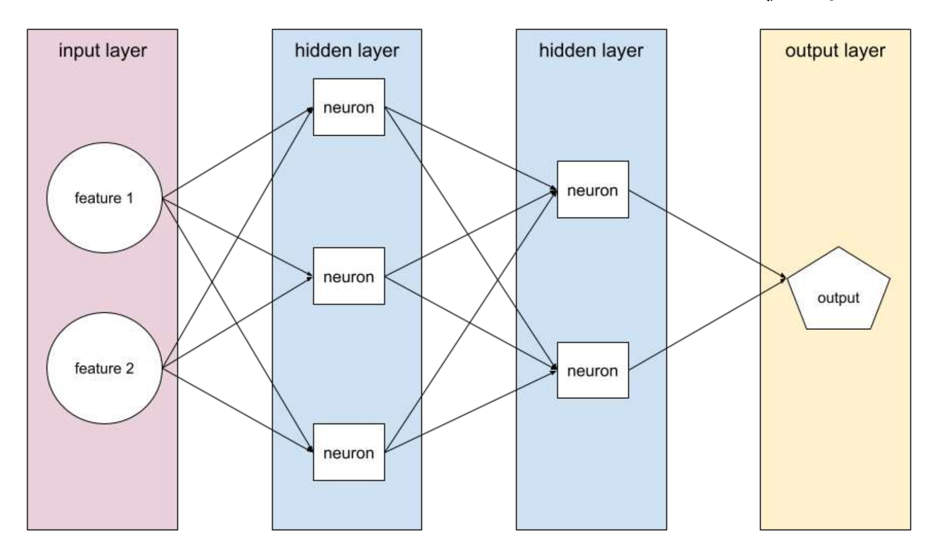
الشبكة العصبية

نموذج (model) يحتوي على طبقة مخفية (hidden layer) واحدة على الأقل. الشبكة العصبية العميقة (deep neural network) هي نوع من الشبكات العصبية التي تحتوي على أكثر من طبقة مخفية. على سبيل المثال، يُظهر الرسم البياني التالي شبكة عصبية عميقة تحتوي على طبقتين مخفيتين.

كل خلية عصبيةفي الشبكة العصبية تتصل بجميع العقدفي الطبقة التالية. على سبيل المثال،في الرسم البياني السابق، لاحظ أن كل خلية من الخلايا العصبية الثلاثةفي الطبقة المخفية الأولى تتصل بشكل منفصل بكل من الخلايا العصبيةفي الطبقة المخفية الثانية.

تسمى الشبكات العصبية المطبقة على أجهزة الكمبيوتر أحيانًا الشبكات العصبية الاصطناعية (artificial neural networks) لتمييزها عن الشبكات العصبية الموجودةفي الدماغ والأنظمة العصبية الأخرى.

يمكن لبعض الشبكات العصبية أن تحاكي العلاقات غير الخطية المعقدة للغاية بين السمات المختلفة والتسمية.



انظر أيضًا الشبكة العصبية التلافيفية (convolutional neural network) والشبكة العصبية المتكررة (recurrent neural network).

Neuron

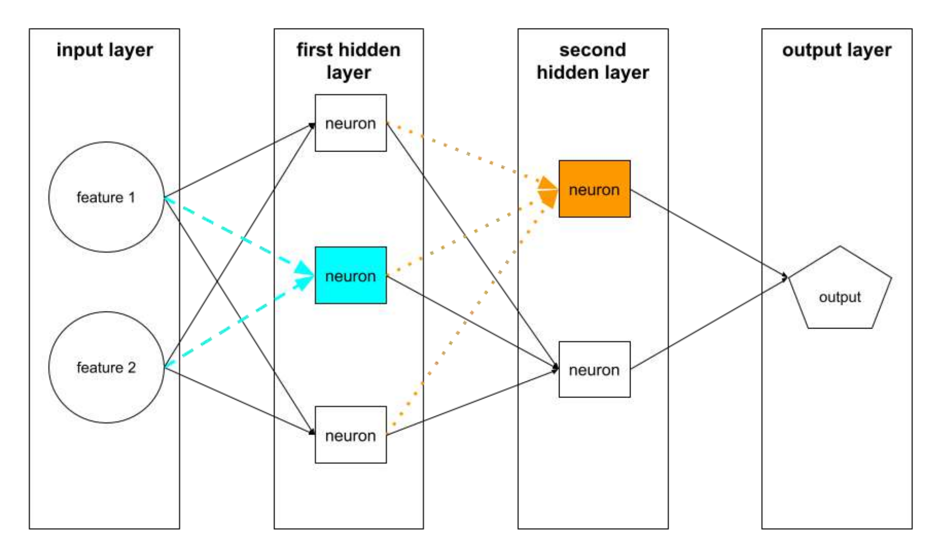
الخلية العصبية

في التعلم الآلي، وحدة متميزة داخل طبقة مخفية (hidden layer) من الشبكة العصبية (neural network). يقوم كل خلية عصبية بتنفيذ الإجراء المكون من خطوتين التاليين:

- لحساب المجموع المرجح (weighted sum) لقيم الإدخال مضروبًافي أوزانها المقابلة.
- يمرر المجموع المرجح كمدخل إلى دالة التنشيط (activation function).

تقبل الخلية العصبية الموجودةفي الطبقة المخفية الأولى المدخلات من قيم الميزاتفي طبقة الإدخال (input layer). تقبل الخلية العصبية الموجودةفي أي طبقة مخفية بعد الطبقة الأولى مدخلات من الخلايا العصبيةفي الطبقة المخفية السابقة. على سبيل المثال، الخلايا العصبيةفي الطبقة المخفية الثانية تقبل المدخلات من الخلايا العصبيةفي الطبقة المخفية الأولى.

يوضح الرسم التوضيحي التالي اثنين من الخلايا العصبية ومدخلاتهما.



تحاكي الخلايا العصبيةفي الشبكة العصبية سلوك الخلايا العصبيةفي الدماغ وأجزاء أخرى من الجهاز العصبي.

NLU

اختصار لفهم اللغة الطبيعية (natural language understanding).

N-gram

N-جرام

هو تسلسل مرتب من كلمات N. على سبيل المثال، "truly madly " هو 2 جرام. لأن الطلب مناسب، "madly truly " هو 2 جرام مختلف عن "truly madly ".

N	Name(s) for this kind of N-gram	Examples
2	bigram or 2-gram	*to go, go to, eat lunch, eat dinner*
3	trigram or 3-gram	*ate too much, three blind mice, the bell tolls*
4	4-gram	*walk in the park, dust in the wind, the boy ate lentils*

تعتمد العديد من نماذج فهم اللغة الطبيعية على N-grams للتنبؤ بالكلمة التالية التي سيكتبها المستخدم أو يقولها. على سبيل المثال، افترض أن مستخدمًا كتب "three blind" . من المحتمل أن يتوقع نموذج NLU المستند إلى trigrams أن المستخدم سوف يكتب تالياً "mice" .

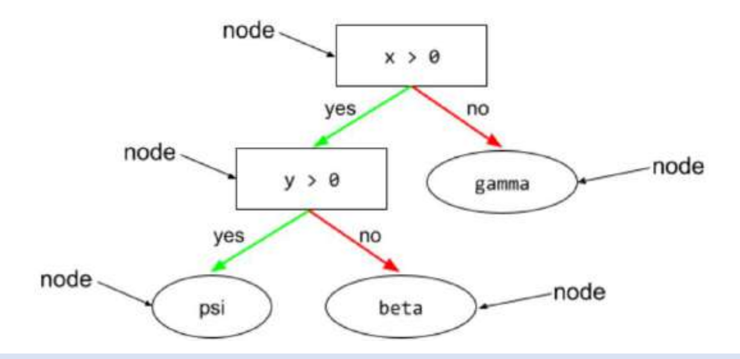
تتباين N-grams مع حقيبة الكلمات (bag of words)، وهي مجموعات غير مرتبة من الكلمات.

Node (neural network)

خلية عصبية (neuron) في طبقة مخفية (hidden layer).

Node (decision tree)

في شجرة القرار (decision tree)، أي شرط (condition) أو ورقة (leaf).



Noise

الضوضاء

أي معلومات غير ذات صلة أو عشوائيةفي مجموعة بيانات تحجب النمط الأساسي.

بشكل عام، أي شيء يحجب الإشارةفي مجموعة البيانات. يمكن إدخال الضوضاءفي البيانات بعدة طرق. فمثلا:

- يخطئ المقيمون البشريونفي وضع العلامات.
- يخطئ البشر والأدواتفي تسجيل قيم السمات أو حذفها.

Nominal Variable

المتغيرات الاسمية

المتغيرات الاسمية هي متغيرات فئوية (categorical variables) لها فئتان أو أكثر دون أي نوع من الترتيب لها.

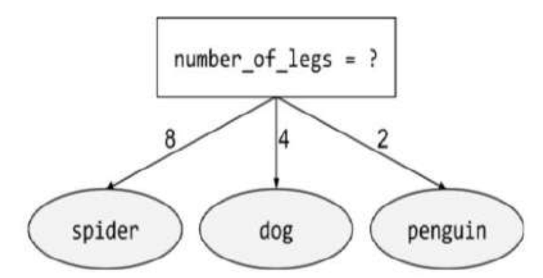
على سبيل المثال، عمود يسمى "اسم المدن" بقيم مثل دلهي وومومباي وتشيناي وما إلى ذلك.

يمكننا أن نرى أنه لا يوجد ترتيب بين المتغيرات ـ بمعنى أن دلهي ليست بأي حال من الأحوال أعلى أو أقل من مومباي (ما لم يذكر صراحة).

Non-binary condition

شرط غير ثنائي

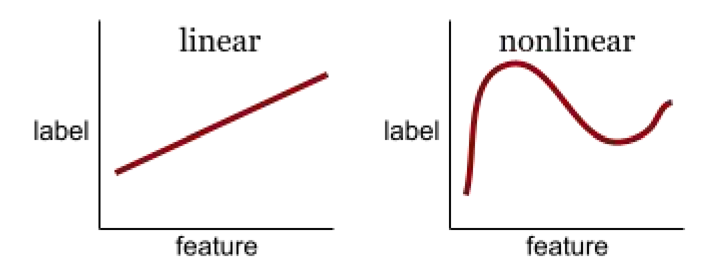
شرط (condition) يحتوي على أكثر من نتيجتين محتملتين. على سبيل المثال، يحتوي الشرط غير الثنائي التالي على ثلاث نتائج محتملة:



Nonlinear

غير خطي

علاقة بين متغيرين أو أكثر لا يمكن تمثيلها من خلال الجمع والضرب فقط. يمكن تمثيل العلاقة الخطية كخط؛ لا يمكن تمثيل العلاقة غير الخطية كخط. على سبيل المثال، ضعفي اعتبارك نموذجين يربط كل منهما ميزة واحدة بتسمية واحدة. النموذج الموجود على اليسار خطي والنموذج الموجود على اليمين غير خطي:



Nonstationarity

عدم الاستقرار

ميزة تتغير قيمها عبر بُعد واحد أو أكثر، عادةً ما يكون الوقت. على سبيل المثال، ضعفي اعتبارك الأمثلة التالية على عدم الاستقرار:

- يختلف عدد ملابس السباحة المباعةفي متجر معين باختلاف الموسم.
- كمية فاكهة معينة يتم حصادهافي منطقة معينة تساوي صفرًافي معظم فترات العام ولكنها كبيرة لفترة وجيزة.
- بسبب تغير المناخ، يتغير متوسط درجات الحرارة السنوية.

على النقيض من الاستقرار (stationarity).

Normal Distribution

التوزيع الطبيعي

التوزيع الطبيعي هو التوزيع الأكثر أهمية والأكثر استخدامًافي الإحصاء. يطلق عليه أحيانًا منحنى الجرس (bell curve)، لأنه يحتوي على شكل غريب من الجرس.في الغالب، يشبه التوزيع الاحتمالي الثنائي (binomial distribution) التوزيع الطبيعي. الفرق بين الاثنين هو التوزيع الطبيعي المستمر.

Normalization

التسوية

بشكل عام، عملية تحويل النطاق الفعلي لقيم المتغير إلى نطاق قياسي من القيم، مثل:

- +1 to -1
- 1 to 0
- التوزيع الطبيعي

على سبيل المثال، افترض أن النطاق الفعلي للقيم لميزة معينة هو 800 إلى 2400. كجزء من هندسة الميزات (feature engineering)، يمكنك تسوية القيم الفعلية وصولاً إلى نطاق قياسي، مثل 1– إلى 1+.

التسوية مهمة شائعةفي هندسة الميزات. عادةً ما تتدرب النماذج بشكل أسرع (وتنتج تنبؤات أفضل) عندما يكون لكل ميزة عدديةفي متجه الميزة (feature vector) نفس النطاق تقريبًا.

Novelty detection

اكتشاف الحداثة

عملية تحديد ما إذاكان مثال جديد (جديد) يأتي من نفس التوزيع مثل مجموعة التدريب. بعبارة أخرى، بعد التدريب على مجموعة التدريب، يحدد اكتشاف الحداثة ما إذا كان المثال الجديد (أثناء الاستدلال أو أثناء التدريب الإضافي) أمرًا متطرفا (outlier).

على النقيض من اكتشاف القيم المتطرفة (outlier detection).

Numerical data

البيانات العددية

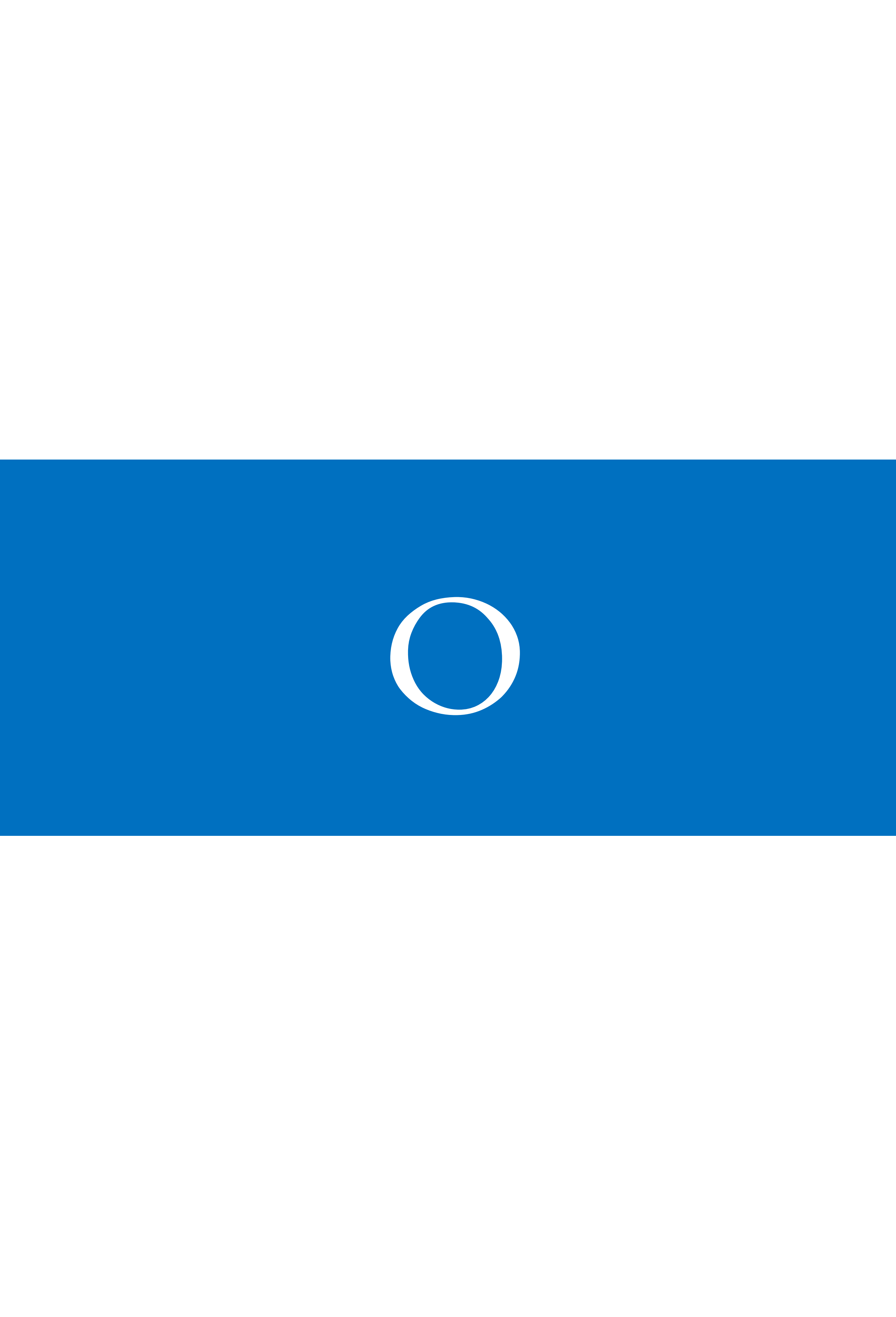
يتم تمثيل الميزات (Features) كأعداد صحيحة أو أرقام حقيقية. على سبيل المثال، من المحتمل أن يمثل نموذج تقييم المنزل حجم المنزل (بالأقدام المربعة أو بالمتر المربع) كبيانات رقمية. يشير تمثيل المعلم على أنه بيانات رقمية إلى أن قيم الميزة لها علاقة رياضية بالتسمية. أي أن عدد الأمتار المربعةفي المنزل ربما يكون له علاقة رياضية بقيمة المنزل.

لا يجب تمثيل جميع بيانات الأعداد الصحيحة كبيانات عددية. على سبيل المثال، تعتبر الرموز البريديةفي بعض أجزاء العالم أعدادًا صحيحة؛ ومع ذلك، لا ينبغي تمثيل الرموز البريدية الصحيحة كبيانات رقميةفي النماذج. هذا لأن الرمز البريدي 20000 ليس ضعف (أو نصف) قوة الرمز البريدي 10000. علاوة على ذلك، على الرغم من أن الرموز البريدية المختلفة ترتبط بقيم عقارات مختلفة، لا يمكننا افتراض أن قيم العقاراتفي الرمز البريدي 20000 هي ضعف قيمة قيم العقارات عند الرمز البريدي 10000. يجب تمثيل الرموز البريدية كبيانات فئوية (categorical data) بدلاً من ذلك.

تسمى الميزات العددية أحيانًا الميزات المستمرة (continuous features).

NumPy

مكتبة رياضية مفتوحة المصدر توفر عمليات مصفوفة فعالةفي بايثون. تم بناء pandas على NumPy.


O

O

Object Detection

اكتشاف الكائنات

اكتشاف الكائنات هي تقنية كمبيوتر تتعلق بالرؤية الحاسوبية (computer vision) ومعالجة الصور (image processing) التي تتعامل مع اكتشاف حالات الكائنات الدلالية لفئة معينةفي الصور ومقاطع الفيديو الرقمية. تشمل المجالات المدروسة جيداً للكشف عن الأشياء اكتشاف الوجه (face detection) واكتشاف المشاة (pedestrian detection).

Objective

الهدف

مقياس تحاول الخوارزمية تحسينه.

Objective function

دالة الهدف

الصيغة الرياضية أو المقياس الذي يهدف النموذج إلى تحسينه. على سبيل المثال، عادةً ما تكون دالة الهدف للانحدار الخطي (linear regression) هي متوسط الخطأ التربيعي (Mean Squared Loss). لذلك، عند تدريب نموذج الانحدار الخطي، يهدف التدريب إلى تقليل (minimize) متوسط الخطأ التربيعي.

في بعض الحالات، يكون الهدف هو تعظيم (maximize) دالة الهدف. على سبيل المثال، إذا كانت دالة الهدف هي الدقة، فإن الهدف هو زيادة الدقة إلى أقصى حد.

انظر أيضا الخطأ (loss).

Oblique condition

شرط مائل

في شجرة القرار (decision tree)، شرط (condition) يتضمن أكثر من ميزة (feature) واحدة. على سبيل المثال، إذا كان الطول والعرض كلاهما ميزتين، فإن ما يلي هو شرط مائل:

```
height > width
```

على النقيض من شرط محاذاة المحور (axis-aligned condition).

Observation

مشاهدة

نقطة بيانات أو صف أو عينةفي مجموعة بيانات.

مرادف لـ (example) و(instance).

Offline

الاوفلاين

مرادف للثابت (static).

Offline inference

الاستدلال الاوفلاين

عملية إنشاء نموذج لمجموعة من التنبؤات ثم تخزين (حفظ) تلك التنبؤات. يمكن للتطبيقات بعد ذلك الوصول إلى التنبؤ المطلوب من ذاكرة التخزين المؤقت بدلاً من إعادة تشغيل النموذج.

على سبيل المثال، ضعفي اعتبارك نموذجًا يولد تنبؤات جوية محلية (تنبؤات) مرة كل أربع ساعات. بعد تشغيل كل نموذج، يخزن النظام مؤقتًا جميع تنبؤات الطقس المحلية. تسترد تطبيقات الطقس التنبؤات من ذاكرة التخزين المؤقت.

يسمى الاستدلال الاوفلاين أيضًا بالاستدلال الثابت (static inference).

على النقيض من الاستدلال الاونلاين (online inference).

One-hot encoding

الترميز واحد-ساخن

تمثل البيانات الفئوية كمتجه حيث:

- تم تعيين عنصر واحد على 1.
- تم تعيين جميع العناصر الأخرى على 0.

يشيع استخدام ترميز واحدــساخن لتمثيل السلاسل أو المعرفات التي تحتوي على مجموعة محدودة من القيم المحتملة. على سبيل المثال، افترض أن ميزة فئوية معينة تسمى "Scandinavia" لها خمس قيم محتملة:

- "Denmark"
- "Sweden"

- "Norway"
- "Finland"
- "Iceland"

يمكن أن يمثل الترميز الواحد_الساخن كل من القيم الخمس على النحو التالي:

country	Vector				
"Denmark"	1	0	0	0	0
"Sweden"	0	1	0	0	0
"Norway"	0	0	1	0	0
"Finland"	0	0	0	1	0
"Iceland"	0	0	0	0	1

بفضل ترميز واحد_ساخن، يمكن للنموذج أن يتعلم اتصالات مختلفة بناءً على كل بلد من البلدان الخمسة.

يمثل تمثيل ميزة كبيانات رقمية (numerical data) بديلاً للترميز الواحد_الساخن. لسوء الحظ، فإن تمثيل الدول الاسكندنافية عدديًا ليس خيارًا جيداً. على سبيل المثال، ضع في اعتبارك التمثيل الرقمي التالي:

- "Denmark" هي 0.
- "Sweden" هي 1.
- "Sweden" هي 2.
- "Sweden" هي 3.
- "Sweden" هي 4.

باستخدام الترميز الرقمي، يفسر النموذج الأرقام الأولية رياضيًا وسيحاول التدرب على هذه الأرقام. ومع ذلك، فإن "Iceland" ليست في الواقع ضعف (أو نصف) من شيء مثل "Norway"، لذا فإن النموذج قد يصل إلى بعض الاستنتاجات الغريبة.

One-shot learning

التعلم دفعة واحدة

نهج التعلم الآلي، الذي يستخدم غالبًا لتصنيف الكائنات، مصمم لتعلم المصنفات الفعالة من مثال تدريب واحد.

انظر أيضا التعلم بلقطة واحدة (one-shot learning).

One-vs.-all

الواحد مقابل الكل

بالنظر إلى مشكلة تصنيف مع فئات N، حل يتكون من N مصنفات ثنائية (binary classifiers) منفصلة ـ مصنف ثنائي واحد لكل نتيجة محتملة. على سبيل المثال، بالنظر إلى نموذج يصنف الأمثلة على أنها حيوانية أو نباتية أو معدنية، فإن حل واحد مقابل الكل سيوفر المصنفات الثنائية الثلاثة المنفصلة التالية:

- حيوان مقابل ليس حيوان.
- الخضروات مقابل ليس خضروات.
- المعدنية مقابل ليس معدنية.

Online

اونلاين

مرادف لديناميكي (dynamic).

Online inference

الاستدلال الاونلاين

توليد التنبؤات (predictions) عند الطلب. على سبيل المثال، افترض أن أحد التطبيقات يمرر إدخالاً إلى نموذج ويصدر طلبًا للتنبؤ. يستجيب النظام الذي يستخدم الاستدلال الاونلاين للطلب عن طريق تشغيل النموذج (وإعادة التنبؤ إلى التطبيق).

على النقيض من الاستدلال الاوفلاين (offline inference).

OpenCv

OpenCV (مكتبة الرؤية الحاسوبية مفتوحة المصدر) هي مكتبة برمجيات الرؤية الحاسوبية (computer vision) والتعلم الآلي (machine learning) مفتوحة المصدر. تم بناء OpenCV لتوفير بنية تحتية مشتركة لتطبيقات الرؤية الحاسوبية ولتسريع استخدام تصور الآلةفي المنتجات التجارية.

Optimizer

المحسن

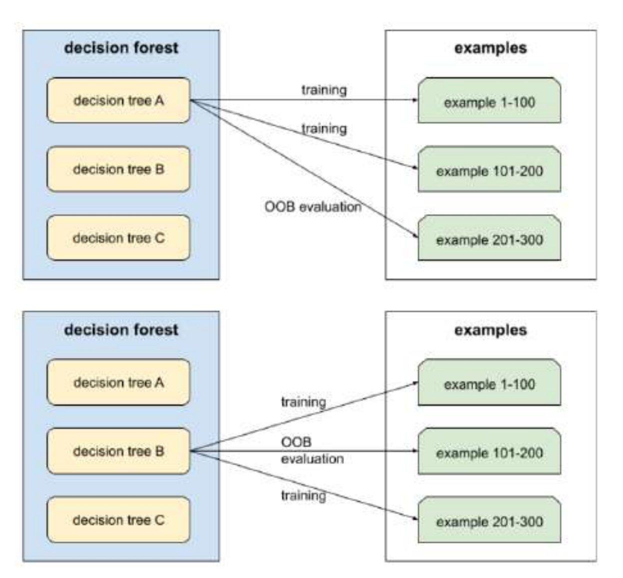
تنفيذ محدد لخوارزمية التدرج الاشتقاقي (gradient descent). تشمل المحسِّنون المشهورون ما يلي:

- AdaGrad ، والتي تعني ADAptive GRADient.
- Adam، والتي تعني ADAptive with Momentum.

Out-of-bag evaluation (OOB evaluation)

التقييم خارج الحقيبة (تقييم OOB)

آلية لتقييم جودة غابة القرارات (decision forest) عن طريق اختبار كل شجرة قرار (decision tree) مقابل الأمثلة (examples) التي لم يتم استخدامها أثناء تدريب (training) شجرة القرار تلك. على سبيل المثال،في الرسم البياني التالي، لاحظ أن النظام يقوم بتدريب كل شجرة قرار على حوالي ثلثي الأمثلة، ثم يقوم بالتقييم مقابل الثلث المتبقي من الأمثلة.



التقييم خارج الحقيبة هو تقريب متحفظ وفعال من الناحية الحسابية لآلية التحقق المتبادل (cross-validation).في التحقق المتبادل، يتم تدريب نموذج واحد لكل جولة تحقق من الصحة المتقاطعة (على سبيل المثال، يتم تدريب 10 نماذج على التحقق المتقاطع 10 أضعاف). مع تقييم OOB، يتم تدريب نموذج واحد. نظرًا لأن التعبئة (bagging) تحجب بعض البيانات من كل شجرة أثناء التدريب، يمكن لتقييم OOB استخدام تلك البيانات لتقريب التحقق المتبادل.

Outlier detection

اكتشاف القيم المتطرفة

عملية تحديد القيم المتطرفة (outlier)في مجموعة التدريب.

على النقيض من اكتشاف الحداثة (novelty detection).

Outliers

القيم المتطرفة

القيم المتطرفة هي قيم بعيدة عن معظم القيم الأخرى.في التعلم الآلي، يعتبر أي مما يلي قيمًا متطرفة:

- بيانات الإدخال التي تزيد قيمها عن 3 انحرافات معيارية عن المتوسط.
- أوزان (Weights) ذات قيم مطلقة عالية.
- القيم المتوقعة بعيدة نسبيًا عن القيم الفعلية.

على سبيل المثال، افترض أن widget-price هو سمة من سمات نموذج معين. افترض أن متوسط widget-price هو 7 يورو مع انحراف معياري قدره 1 يورو. وبالتالي، فإن الأمثلة التي تحتوي على widget-price 12 يورو أو 2 يورو تعتبر قيمًا متطرفة لأن كل من هذه الأسعار هي خمس انحرافات معيارية عن المتوسط.

غالبًا ما تحدث القيم المتطرفة بسبب الأخطاء المطبعية أو أخطاء الإدخال الأخرى.في حالات أخرى، القيم المتطرفة ليست أخطاء؛ بعد كل شيء، قيم خمس انحرافات معيارية بعيدة عن المتوسط نادرة ولكنها ليست مستحيلة.

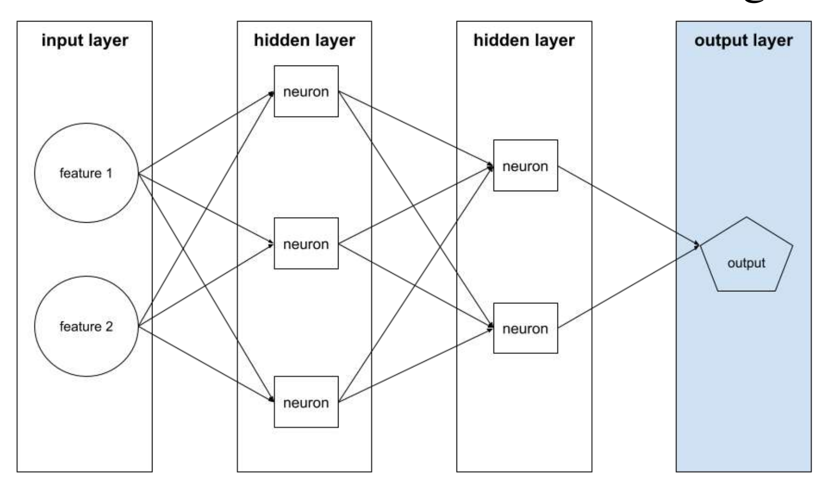
غالبًا ما تسبب القيم المتطرفة مشاكل في التدريب النموذجي. القص (Clipping) هي إحدى طرق إدارة القيم المتطرفة.

Output layer

طبقة الإخراج

الطبقة "النهائية" للشبكة العصبية. تحتوي طبقة الإخراج على التنبؤ.

يوضح الرسم التوضيحي التالي شبكة عصبية صغيرة عميقة بطبقة إدخال وطبقتين مخفيتين وطبقة إخراج:



Overfitting

الضبط الزائد (الملائمة الزائدة)

إنشاء نموذج يطابق بيانات التدريب بشكل وثيق بحيث يفشل النموذجفي إجراء تنبؤات صحيحة على البيانات الجديدة.

يمكن أن يقلل التنظيم (Regularization) من الضبط الزائد. يمكن أن يؤدي التدريب على مجموعة تدريب كبيرة ومتنوعة أيضًا إلى تقليل الضبط الزائد.

يشبه الضبط الزائد بشكل صارم اتباع نصيحة معلمك المفضل فقط. من المحتمل أن تكون ناجحًافي فصل هذا المعلم، ولكن قد "تتأقلم" مع أفكار هذا المعلم وتكون غير ناجحفي الفصول الأخرى. سيمكنك اتباع نصيحة خليط من المعلمين من التكيف بشكل أفضل مع المواقف الجديدة.

Oversampling

الإفراط في أخذ العينات

إعادة استخدام أمثلة (examples) فئة الأقلية (minority class)في مجموعة بيانات غير متوازنة في الفئة (class-imbalanced dataset) من أجل إنشاء مجموعة تدريب أكثر توازناً.

على سبيل المثال، ضعفي اعتبارك مشكلة تصنيف ثنائية (binary classification) تكون فيها نسبة فئة الأغلبية (majority class) إلى فئة الأقلية 5,000:1. إذا كانت مجموعة البيانات تحتوي على مليون مثال، فإن مجموعة البيانات لا تحتوي إلا على حوالي 200 مثال لفئة الأقلية، والتي

قد تكون أمثلة قليلة جداً للتدريب الفعال. للتغلب على هذا النقص، يمكنك الإفراط في أخذ عينات (إعادة استخدام) هذه 200 مثال عدة مرات، وربما تسفر عن أمثلة كافية للتدريب المفيد.

تحتاج إلى توخي الحذر بشأن الضبط الزائد (overfitting) عند الإفراط في أخذ العينات.

على النقيض من نقص العينات (undersampling).

P

P

Padding

الحشو

الحشو هو مصطلح ذو صلة بالشبكات العصبية التلافيفية (CNN) حيث يشير إلى كمية البكسل المضافة إلى الصورة عند معالجتها بواسطة نواة CNN.

عملية إضافة قيمة (غالبا صفر) إلى حواف الصورة، اما لتعديل ابعاد الصورة بحيث تتداخل مع ابعاد مدخلات الشبكة او للحفاظ على ابعاد الصورة قبل تطبيق عملية الالتفاف (convolution operation) وبعده؛ لان تطبيق هذا العملية يؤدي الى خفض الابعاد.

هناك أنواع من الحشو: (Same padding) و (Causal padding) و(Valid padding).

Pandas

واجهة برمجة تطبيقات (API) لتحليل البيانات الموجه نحو العمود مبنية على (numpy). تدعم العديد من أطر التعلم الآلي، بمافي ذلك (TensorFlow)، هياكل بيانات pandas كمدخلات.

Parameter

المعامل

الأوزان (weights) والتحيزات (biases) التي يتعلمها النموذج أثناء التدريب (training). على سبيل المثال،في نموذج الانحدار الخطي (linear regression)، تتكون المعلمات من الانحياز (b) وجميع الأوزان (w1، w2 ، وما إلى ذلك)في الصيغة التالية:

$$y' = b + w_1x_1 + w_2x_2 + \ldots w_nx_n$$

في المقابل، تعد المعلمة الفائقة (hyperparameter) هي القيم التي تقدمها (أو خدمة تحويل المعلمة الفائقة) إلى النموذج. على سبيل المثال، معدل التعلم (learning rate) هو معلمة فائقة.

تتضمن أمثلة المعلمات ما يلي:

- الأوزان (weights)في شبكة عصبية اصطناعية (neural network).
- متجهات الدعم (support vectors)في آلة متجهات الدعم (SVM).
- المعاملات (coefficients)في الانحدار الخطي (linear regression) أو اللوجستي (logestic regression).

Parameter update

تحديث المعامل

عملية تعديل معلمات (parameters) النموذج أثناء التدريب، وعادةً ما يتم ذلك خلال تكرار واحد من التدرج العشوائي (gradient descent).

Partial derivative

الاشتقاق الجزئية

مشتق تعتبر فيه جميع المتغيرات باستثناء واحد منها ثابتًا. على سبيل المثال، المشتق الجزئي لـ *f(x, y)* بالنسبة إلى x هو مشتق f الذي يعتبر دالة لـ x وحده (أي ، الحفاظ على y ثابتًا). يركز المشتق الجزئي لـ f بالنسبة إلى x فقط على كيفية تغير x ويتجاهل جميع المتغيرات الأخرىفي المعادلة.

Pattern Recognition

تمييز الأنماط

تمييز الأنماط هو فرع من فروع التعلم الآلي الذي يركز على التعرف على الأنماط والانتظامفي البيانات. التصنيف (Classification) هو مثال على التعرف على الأنماط حيث يتم تعيين كل قيمة إدخال واحدة من مجموعة معينة من الفئات.

في الرؤية الحاسوبية (computer vision)، تُستخدم تقنيات التعرف على الأنماط الخاضعة للإشراف للتعرف الضوئي على الأحرف (OCR)، واكتشاف الوجوه (face detection)، والتعرف على الوجوه (face recognition)، واكتشاف الأشياء (object detection)، وتصنيف الكائنات (object classification).

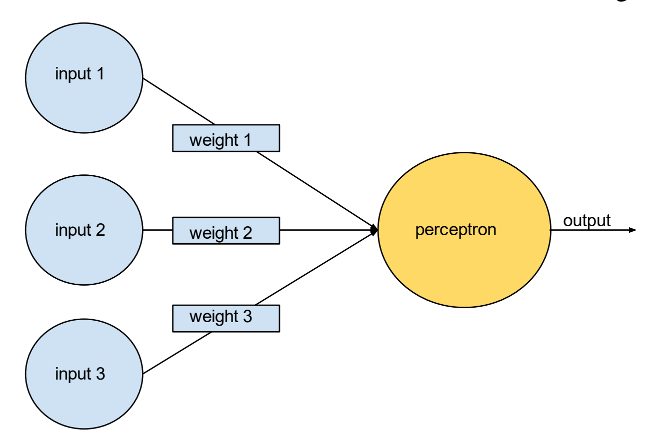
Perceptron

بيرسيبترون

نظام (سواء كان جهازًا أو برنامجًا) يأخذ قيمة إدخال واحدة أو أكثر، ويقوم بتشغيل دالة على المجموع المرجح للمدخلات، ويحسب قيمة إخراج واحدة.في التعلم الآلي، تكون الدالة عادةً غير خطية، مثل ReLU أو sigmoid أو tanh. على سبيل المثال، يعتمد البيرسيبترون التالي على دالة sigmoid لمعالجة ثلاث قيم إدخال:

$$f(x_1, x_2, x_3) = \text{sigmoid}(w_1x_1 + w_2x_2 + w_3x_3)$$

في الرسم التوضيحي التالي، يأخذ البيرسيبترون ثلاث مدخلات، كل منها يتم تعديله بنفسه بواسطة الوزن قبل دخول البيرسيبترون:



البيرسيبترون هي الخلايا العصبية (neurons)في الشبكات العصبية (neural networks).

Permutation variable importance

تباديل أهمية المتغير

نوع من أهمية المتغير (variable importance) يقوم بتقييم الزيادةفي خطأ التنبؤ بنموذج بعد تبديل قيم الميزة. تقليب أهمية متغير هو مقياس نموذج حيادي.

Perplexity

الارتباك

مقياس واحد لمدى نجاح النموذجفي إنجاز مهمته. على سبيل المثال، افترض أن مهمتك هي قراءة الأحرف القليلة الأولى من كلمة يكتبها المستخدم على لوحة مفاتيح هاتف ذكي، وتقديم قائمة بكلمات الإكمال المحتملة. الارتباك، P، لهذه المهمة هي تقريبًا عدد التخمينات التي تحتاج إلى تقديمها حتى تحتوي قائمتك على الكلمة الفعلية التي يحاول المستخدم كتابتها.

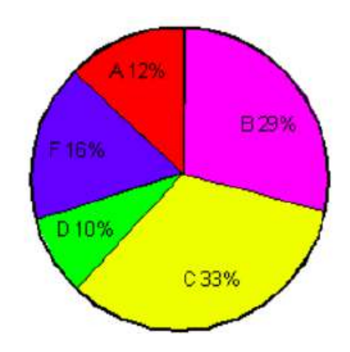
يرتبط الارتباك بالانتروبيا المتبادلة(cross-entropy) على النحو التالي:

$$P = 2^{-\text{cross entropy}}$$

Pie Chart

المخطط الدائري

المخطط الدائري هو رسم إحصائي دائري مقسم إلى شرائح لتوضيح النسبة العددية. يتناسب طول القوس لكل شريحة مع الكمية التي تمثلها. دعونا نفهمها بمثال:



يمثل هذا رسمًا بيانيًا دائريًا يوضح نتائج الاختبار. يتم الإشارة إلى كل درجة بواسطة "شريحة". مجموع النسب المئوية يساوي 100. مجموع قياسات القوس يساوي 360 درجة. إذن، حصل 12٪ من الطلاب على درجة A و29٪ حصلوا على درجة B وهكذا.

Pipeline

مسارات

البنية التحتية المحيطة بخوارزمية التعلم الآلي. يتضمن المسارات جمع البيانات، ووضع البيانات في ملفات بيانات التدريب، وتدريب نموذج واحد أو أكثر، وتصدير النماذج إلى الإنتاج.

Pipelining

شكل من أشكال التوازي للنموذج (model parallelism) يتم فيه تقسيم معالجة النموذج إلى مراحل متتالية ويتم تنفيذ كل مرحلة على جهاز مختلف. أثناء معالجة المرحلة لدفعة واحدة، يمكن أن تعمل المرحلة السابقة على الدُفعة التالية.

انظر أيضا التدريب على مراحل (staged training).

Policy

السياسة

في التعلم المعزز (reinforcement learning)، رسم الخرائط الاحتمالية للوكيل (agent) من الحالات (states) إلى الأفعال (actions).

Polynomial Regression

الانحدار متعدد الحدود

الانحدار متعدد الحدود هو شكل من أشكال الانحدار الخطي (Linear regression) المعروف باسم حالة خاصة من الانحدار الخطي المتعدد (Multiple linear regression) الذي يقدر العلاقة على أنها متعددة الحدود من الدرجة n. يعد الانحدار متعدد الحدود حساسًا للقيم

المتطرفة (outliers)، لذا فإن وجود واحد أو اثنين من القيم المتطرفة يمكن أن يؤثر أيضًا بشكل سيء على الأداء.

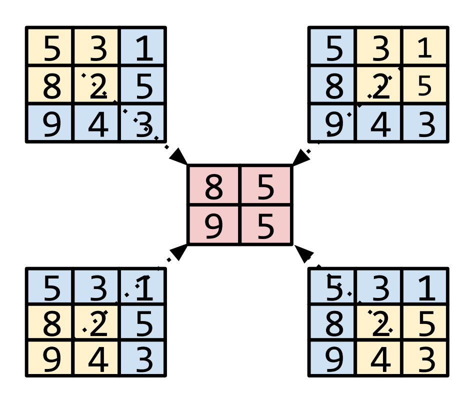
Pooling

التجميع

اختزال مصفوفة (أو مصفوفات) تم إنشاؤها بواسطة طبقة تلافيفية (convolutional layer) سابقة إلى مصفوفة أصغر. عادةً ما يتضمن التجميع أخذ القيمة القصوى (max pooling) أو المتوسطة (average pooling) عبر المنطقة المُجمَّعة. على سبيل المثال، افترض أن لدينا مصفوفة x33 التالية:

5	3	1
8	2	5
9	4	3

عملية التجميع، تمامًا مثل العملية التلافيفية، تقسم تلك المصفوفة إلى شرائح ثم تنزلق هذه العملية التلافيفية بخطوات (strides). على سبيل المثال، افترض أن عملية التجميع قسمت المصفوفة التلافيفية إلى شرائح 2 × 2 بخطوة 1 × 1. كما يوضح الرسم البياني التالي، تتم أربع عمليات تجميع. تخيل أن كل عملية تجميع تختار الحد الأقصى لقيمة الأربعةفي تلك الشريحة:



يساعد التجميع على فرض الثبات الترجمي (translational invariance)في مصفوفة الإدخال.

يُعرف التجميع لتطبيقات الرؤية بشكل أكثر رسمية باسم التجميع المكاني (spatial pooling). تشير تطبيقات السلاسل الزمنية عادةً إلى التجميع على أنه تجميع زماني (temporal pooling).

بشكل أقل رسمية، يُطلق على التجميع غالبًا الاختزال الجزئي (subsampling) أو الاختزال (downsampling).

Positive class

الفئة الإيجابية

الصف الذي تختبر له.

على سبيل المثال، قد تكون الفئة الإيجابيةفي نموذج السرطان هي "الورم". قد تكون الفئة الإيجابية في مصنف البريد الإلكتروني هي "البريد العشوائي".

على النقيض من الفئة السلبية (negative class).

PR AUC (area under the PR curve)

المنطقة الواقعة تحت منحنى الدقةـ الاسترجاع (precision-recall curve) المحرف، التي تم الحصول عليها بالتخطيط (الاسترجاع (recall) والدقة (precision)) للنقاط لقيم مختلفة لعتبة التصنيف (classification threshold). اعتمادًا على كيفية حسابها، قد تكون PR AUC معادلة لمتوسط دقة (average precision) النموذج.

Precision

الدقة

مقياس لنماذج التصنيف يجيب على السؤال التالي:

عندما تنبأ النموذج بالفئة الإيجابية، ما هي النسبة المئوية الصحيحة للتنبؤات؟

ها هي الصيغة:

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

حيث:

- الإيجابي الحقيقي (true positive) يعني أن النموذج تنبأ بالفئة الإيجابية بشكل صحيح.
- الإيجابي الخاطئ (false positive) يعني أن النموذج تنبأ عن طريق الخطأ بالفئة الإيجابية.

على سبيل المثال، افترض أن نموذجًا قدم 200 تنبؤ إيجابي. من بين هذه التوقعات الإيجابية البالغ عددها 200:

- 150 كانت إيجابية حقيقية.
- 50 كانت ايجابيات خاطئة.

في هذه الحالة:

$$\text{Precision} = \frac{150}{150 + 50} = 0.75$$

على النقيض من الدقة (accuracy) والاسترجاع (recall).

Precision-recall curve

منحنى الدقة-الاسترجاع

منحنى الدقة (precision) مقابل الاسترجاع (recall) عند عتبات تصنيف (classification thresholds) مختلفة.

Prediction

التنبؤ

ناتج النموذج. فمثلا:

- التنبؤ بنموذج التصنيف الثنائي هو إما الفئة الموجبة أو الفئة السلبية.
- التنبؤ بنموذج التصنيف متعدد الفئات هو فئة واحدة.
- التنبؤ بنموذج الانحدار الخطي هو رقم.

Prediction bias

تحيز التنبؤ

قيمة تشير إلى مدى تباعد متوسط التنبؤات (predictions) عن متوسط التسمياتفي مجموعة البيانات.

Pre-trained model

نموذج المدرب مسبقا

النماذج أو مكونات النموذج (مثل متجه التضمين (embedding vector)) التي تم تدريبها بالفعل.في بعض الأحيان، ستقوم بتغذية متجهات التضمين المدربة مسبقًافي شبكة عصبية (neural network).في أوقات أخرى، سيقوم نموذجك بتدريب متجهات التضمين نفسها بدلاً من الاعتماد على التضمينات المدربة مسبقًا.

Principal Component Analysis (PCA)

تحليل المكونات الأساسيّة

يعد تحليل المكون الرئيسي (PCA) أسلوبًا شائعًا لتحليل مجموعات البيانات الكبيرة التي تحتوي على عدد كبير من الأبعاد (dimensions) / الميزات (features) لكل ملاحظة، مما يزيد من قابلية تفسير البيانات مع الحفاظ على أكبر قدر من المعلومات، وتمكين تصور البيانات متعددة الأبعاد. بشكل رسمي، تعد PCA تقنية إحصائية لتقليل أبعاد مجموعة البيانات.

Prior belief

الاعتقاد المسبق

ما تعتقده بشأن البيانات قبل أن تبدأ التدريب عليها. على سبيل المثال، يعتمد تنظيم L_2 (L_2 regularization) على اعتقاد مسبق بأن الأوزان يجب أن تكون صغيرة وموزعة بشكل طبيعي حول الصفر.

Probabilistic regression model

نموذج الانحدار الاحتمالي

نموذج انحدار (regression model) لا يستخدم فقط أوزان (weights) كل سمة (feature)، بل يستخدم أيضًا عدم اليقين في تلك الأوزان. يولد نموذج الانحدار الاحتمالي تنبوًأ وعدم اليقين من هذا التنبؤ. على سبيل المثال، قد ينتج عن نموذج الانحدار الاحتمالي تنبؤ قدره 325 مع انحراف معياري قدره 12.

Proxy labels

تسميات الوكيل

البيانات المستخدمة لتقريب التسميات غير المتوفرة مباشرةفي مجموعة البيانات.

على سبيل المثال، افترض أنه يجب عليك تدريب نموذج للتنبؤ بمستوى إجهاد الموظف. تحتوي مجموعة البيانات الخاصة بك على الكثير من الميزات التنبؤية ولكنها لا تحتوي على تصنيف يسمى مستوى الإجهاد. وبكل شجاعة، يمكنك اختيار "حوادث مكان العمل" كتسمية وكيل لمستوى التوتر. بعد كل شيء، يتعرض الموظفون الذين يتعرضون لضغوط شديدة إلى حوادث أكثر من الموظفين الهادئين. أم هم؟ ربما ترتفع حوادث مكان العمل وتنخفض لأسباب متعددة.

كمثال ثان، افترض أنك تريد هل تمطر؟ أن تكون تسمية منطقية لمجموعة البيانات الخاصة بك، ولكن لا تحتوي مجموعة البيانات الخاصة بك على بيانات المطر.في حالة توفر الصور، يمكنك إنشاء صور لأشخاص يحملون مظلات كعلامة وكيل هل تمطر؟ هل هذه تسمية وكيل جيدة؟

ربما، ولكن قد يكون الناس في بعض الثقافات أكثر عرضة لحمل المظلات للحماية من أشعة الشمس أكثر من المطر.

غالبًا ما تكون تسميات الوكيل غير كاملة. عندما يكون ذلك ممكنًا، اختر تسميات فعلية على تسميات الوكيل. ومع ذلك،في حالة عدم وجود تسمية فعلية، اختر تسمية الوكيل بعناية شديدة، واختر مرشح تسمية الوكيل الأقل فظاعة.

Pruning

التقليم

التقليم هو تقنية تزيل أجزاء شجرة القرار (decision tree) التي تمنعها من النمو إلى عمقها الكامل. الأجزاء التي تزيلها من الشجرة هي الأجزاء التي لا توفر القدرة على تصنيف المثيلات. الهدف من التقليم هو إزالة الفروع غير المرغوب فيها وتحسين بُنية الشجرة.

P-Value

قيمة P

P-value هي قيمة احتمال الحصول على نتيجة مساوية للقيمة المرصودة أو أكبر منها ، عندما تكون الفرضية الصفرية صحيحة.

Python

بايثون

بايثون هي لغة برمجة مفتوحة المصدر، تُستخدم على نطاق واسع في العديد من التطبيقات، مثل البرمجة للأغراض العامة وعلم البيانات والتعلم الآلي. يفضل عادة المبتدئين في هذه المجالات بسبب المزايا الرئيسية التالية:

- سهلة التعلم.
- لغة عالية المستوى.
- تستخدم على نطاق واسع ومدعومة.

Pytorch

PyTorch هو إطار عمل مفتوح المصدر للتعلم الآلي (ML) يعتمد على لغة برمجة Python ومكتبة Torch. إنها واحدة من المنصات المفضلة لأبحاث التعلم العميق. تم تصميم إطار العمل لتسريع العملية بين النماذج الأولية للبحث والنشر. يستخدم لتطبيقات مثل الرؤية الحاسوبية (computer vision) ومعالجة اللغة الطبيعية (Natural Language Processing)، تم تطويره

في الأصل بواسطة Meta AI وهو الآن جزء من مظلة Linux Foundation. إنه برنامج مجاني ومفتوح المصدر تم طرحه بموجب ترخيص BSD المعدل.

Q

Q

Q-function

دالة-Q

في التعلم المعزز (reinforcement learning)، الدالة التي تتوقع العائد (return) المتوقع من اتخاذ إجراء (action) في حالة (state) ما ثم اتباع سياسة (policy) معينة.

تُعرف دالة Q أيضًا باسم دالة قيمة إجراء_الحالة (state-action value function).

Q-learning

تعلم-Q

في التعلم المعزز (reinforcement learning)، خوارزمية تسمح للوكيل (agent) بتعلم دالة_Q (Q-function) المثلى لعملية قرار ماركوف (Markov decision process) من خلال تطبيق معادلة بيلمان (Bellman equation). تشكل عملية قرار ماركوف بيئة.

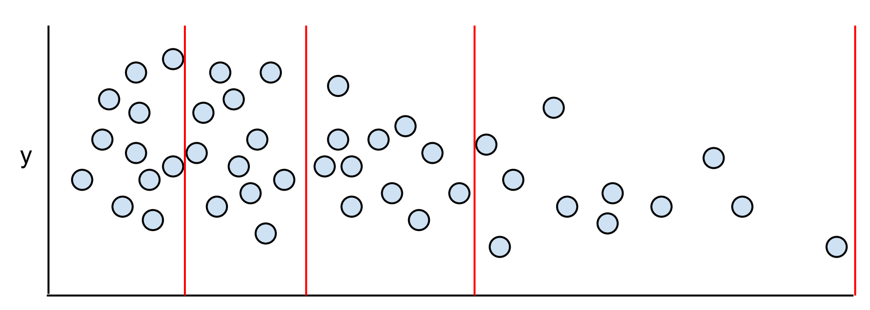
Quantile

كل مجموعة (دلو) في التجميع الكمي (quantile bucketing).

Quantile bucketing

التجميع الكمي

توزيع قيم الميزات في مجموعات (buckets) بحيث تحتوي كل مجموعة (دلو bucket) على نفس (أو نفس عدد الأمثلة تقريبًا). على سبيل المثال، يقسم الشكل التالي 44 نقطة إلى 4 مجموعات، يحتوي كل منها على 11 نقطة. لكي تحتوي كل مجموعة في الشكل على نفس عدد النقاط، تمتد بعض المجموعات على عرض مختلف لقيم x.

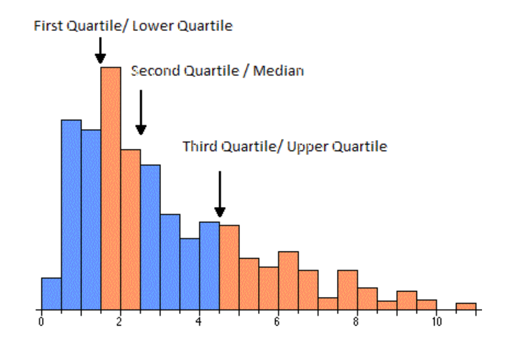


Quartile

الربعي

الربعي يقسم سلسلة إلى 4 أجزاء متساوية. بالنسبة لأي سلسلة، هناك أربعة أرباع يُشار إليها بالرموز Q1و Q2 و Q3 و .Q4 تُعرف هذه بالربع الأول والربع الثاني وما إلى ذلك.

على سبيل المثال، يوضح الرسم البياني أدناه الدرجة الصحية لمريض من النطاق 0 إلى 60. تقسم الأرباع السكان إلى 4 مجموعات.



Quantization

التكميم

خوارزمية تنفذ التجميع الكمي (quantile bucketing) على ميزة (feature) معينةفي مجموعة البيانات (dataset).

R

R

Random forest

الغابة العشوائية

مجموعة (ensemble) من أشجار القرار (decision trees) يتم فيها تدريب كل شجرة قرار بضوضاء عشوائية محددة، مثل التعبئة (bagging).

الغابات العشوائية هي نوع من غابات القرار (decision forest).

Random policy

السياسة العشوائية

في التعلم المعزز (reinforcement learning)، سياسة (policy) تختار الإجراء (action) بشكل عشوائي.

Ranking

الترتيب

نوع من التعلم الخاضع للإشراف (supervised learning) هدفه ترتيب قائمة بالعناصر.

Rank (ordinality)

الرتبة (ترتيب)

الموضع الترتيبي للفئةفي مشكلة التعلم الآلي الذي يصنف الفئات من الأعلى إلى الأدنى. على سبيل المثال، يمكن لنظام تصنيف السلوك أن يصنف مكافآت الكلب من الأعلى (شريحة لحم) إلى الأقل (كرنب ذابلة).

Rater

المقيم

إنسان يقدم تسميات (labels) لأمثلة (examples). "Annotator" هو اسم آخر للمقيم.

Recall

الاسترجاع

مقياس لنماذج التصنيف (classification models) يجيب على السؤال التالي:

عندما كانت الحقيقة الأساسية (ground truth) هي الفئة الإيجابية (positive class)، ما هي النسبة المئوية للتنبؤات التي حددها النموذج بشكل صحيح على أنها فئة إيجابية؟

حيث الصيغة هي:

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

حيث:

- الإيجابي الحقيقي (true positive) يعني أن النموذج تنبأ بالفئة الإيجابية بشكل صحيح.
- تعني السلبية الخاطئة (false negative) أن النموذج تنبأ عن طريق الخطأ بالفئة السلبية (negative class).

على سبيل المثال، افترض أن نموذجك قد أجرى 200 تنبؤ على أمثلة كانت الحقيقة الأساسية بالنسبة لها هي الفئة الإيجابية. من بين 200 تنبؤ:

- 180 كانت إيجابية حقيقية.
- 20 كانت سلبيات كاذبة.

في هذه الحالة:

$$\text{Recall} = \frac{180}{180 + 20} = 0.9$$

يعد الاسترجاع مفيداً بشكل خاص في تحديد القدرة التنبؤية لنماذج التصنيف التي تكون فيها الفئة الإيجابية نادرة. على سبيل المثال، ضع في اعتبارك مجموعة بيانات غير متوازنة في الفئة (class-imbalanced dataset) تحدث فيها الفئة الإيجابية لمرض معين في 10 مرضى فقط من بين مليون. افترض أن نموذجك يقوم بخمسة ملايين توقع ينتج عنه النتائج التالية:

- 30 ايجابيات حقيقية.
- 20 سلبيات كاذبة.
- 4،999،000 سلبية حقيقية.
- 950 ايجابيات كاذبة.

لذلك فإن الاسترجاع لهذا النموذج هو:

```
recall = TP / (TP + FN)
recall = 30 / (30 + 20) = 0.6 = 60%
```

على النقيض من ذلك، فإن دقة (accuracy) هذا النموذج هي:

```
accuracy = (TP + TN) / (TP + TN + FP + FN)
accuracy = (30 + 4,999,000) / (30 + 4,999,000 + 950 +
20) = 99.98%
```

تبدو هذه القيمة العالية للدقة مثيرة للإعجاب ولكنهافي الأساس لا معنى لها. يعد الاسترجاع مقياسًا أكثر فائدة لمجموعات البيانات غير المتوازنةفي الفئة مقارنة بالدقة.

Recommendation Engine

محرك التوصية

يميل الناس عمومًا إلى شراء المنتجات التي أوصى بها أصدقاؤهم أو الأشخاص الذين يثقون بهم.في الوقت الحاضرفي العصر الرقمي، يستخدم أي متجر عبر الإنترنت تقوم بزيارته نوعًا من محرك التوصية. محركات التوصية هيفي الأساس أدوات تصفية البيانات التي تستخدم الخوارزميات والبيانات للتوصية بالعناصر الأكثر صلة بمستخدم معين. إذا تمكنا من التوصية بعناصر للعميل بناءً على احتياجاته واهتماماته، فسيؤدي ذلك إلى إحداث تأثير إيجابي على تجربة المستخدم وسيزوره بشكل متكرر. هناك أنواع قليلة من محركات التوصية:

- التصفية على أساس المحتوى (Content based filtering).
- التصفية التعاونية (Collaborative filtering).
  - تصفية تعاونية بين المستخدم والمستخدم.
  - التصفية التعاونية للعنصر والعنصر.
- أنظمة التوصية الهجينة.

Recommendation system

نظام التوصية

نظام يختار لكل مستخدم مجموعة صغيرة نسبيًا من العناصر (items) المرغوبة من مجموعة كبيرة. على سبيل المثال، قد يوصي نظام توصية الفيديو بمقطعي فيديو من مجموعة من 100000 مقطع فيديو، واختيار Casablanca و The Philadelphia Story لمستخدم واحد، و Wonder Woman و Black Panther لمستخدم آخر. قد يبني نظام توصية الفيديو توصياته على عوامل مثل:

- الأفلام التي قام مستخدمون مشابهون بتصنيفها أو مشاهدتها.
- النوع والمخرجين والممثلين والهدف الديموغرافي .

Rectified Linear Unit (ReLU)

الوحدة الخطية المصححة (ReLU)

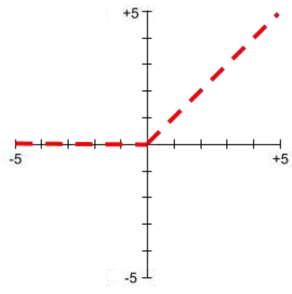
دالة تنشيط (activation function) بالسلوك التالي:

- إذا كان الإدخال سالبًا أو صفرًا، فسيكون الناتج 0.
- إذا كان الإدخال موجبًا، فسيكون الناتج مساويًا للإدخال.

فمثلا:

- إذا كان الإدخال 3ــ، فسيكون الناتج 0.
- إذا كان الإدخال 3+، فإن الناتج يكون 3.0.

هنا مخطط ReLU:



ReLU هي دالة تنشيط شائعة جدًا. على الرغم من سلوكه البسيط، لا يزال ReLU يمكّن الشبكة العصبية من تعلم العلاقات غير الخطية (nonlinear) بين الميزات (features) والتسمية (label).

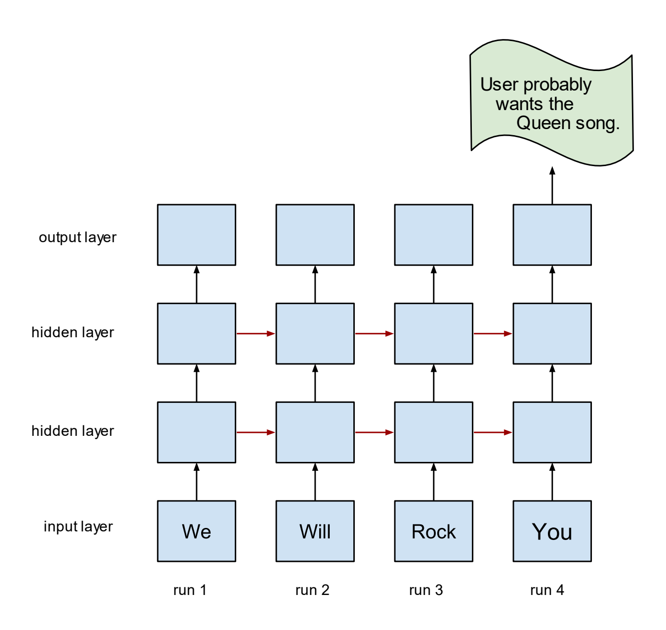
Recurrent neural networks (RNN)

الشبكات العصبية المتكررة

شبكة عصبية (neural network) يتم تشغيلها عن قصد عدة مرات، حيث يتم تغذية أجزاء من كل شوطفي الجولة التالية. على وجه التحديد، توفر الطبقات المخفية من التشغيل السابق جزءًا من الإدخال إلى نفس الطبقة المخفيةفي التشغيل التالي. تعد الشبكات العصبية المتكررة (RNN) مفيدة بشكل خاصفي تقييم التسلسلات (sequences)، بحيث يمكن للطبقات المخفية التعلم من عمليات التشغيل السابقة للشبكة العصبيةفي الأجزاء السابقة من التسلسل.

على سبيل المثال، يوضح الشكل التالي شبكة عصبية متكررة تعمل أربع مرات. لاحظ أن القيم التي تم تعلمهافي الطبقات المخفية من التشغيل الأول أصبحت جزءًا من الإدخال إلى نفس الطبقات المخفيةفي التشغيل الثاني. وبالمثل، فإن القيم التي تم تعلمهافي الطبقة المخفيةفي التشغيل الثاني تصبح جزءًا من الإدخال إلى نفس الطبقة المخفيةفي التشغيل الثالث. وبهذه

الطريقة، تدرب الشبكة العصبية المتكررة وتتنبأ تدريجياً بمعنى التسلسل بأكمله بدلاً من مجرد معنى الكلمات الفردية.



Region-Based Convolutional Neural Network (R-CNN)

كان الهدف الأصلي لـ R-CNN هو التقاط صورة إدخال وإنتاج مجموعة من المربعات المحيطة (bounding boxes) كإخراج، حيث يحتوي كل مربع محيط على كائن وكذلك الفئة (مثل السيارة أو المشاة) للكائن.في الآونة الأخيرة، تم تمديد R-CNN لأداء مهام الرؤية الحاسوبية الأخرى. يغطي ما يلي بعض إصدارات R-CNN التي تم تطويرها.

- R-CNN
- Fast R-CNN
- Faster R-CNN
- Mask R-CNN
- Mesh R-CNN

Regression model

نموذج الانحدار

بشكل غير رسمي، نموذج يولد تنبوًأ عدديًا. (في المقابل، يُنشئ نموذج التصنيف (classification model) تنبوًأ بالفئة) على سبيل المثال، فيما يلي جميع نماذج الانحدار:

- نموذج يتنبأ بقيمة منزل معين، مثل 423000 يورو.

- نموذج يتنبأ بمتوسط العمر المتوقع لشجرة معينة، مثل 23.2 سنة.
- نموذج يتنبأ بكمية الأمطار التي ستسقط في مدينة معينة خلال الساعات الست القادمة، مثل 0.18 بوصة.

هناك نوعان شائعان من نماذج الانحدار هما:

- الانحدار الخطي (Linear regression)، الذي يبحث عن الخط الذي يناسب قيم التسمية بشكل أفضل للميزات.
- الانحدار اللوجستي (Logistic regression)، والذي يولد احتمالًا بين 0.0 و 1.0 يقوم النظام عادةً بتعيينه لتنبؤ الفئة.

ليس كل نموذج ينتج تنبؤات عددية هو نموذج انحدار.في بعض الحالات، يكون التنبؤ الرقمي مجرد نموذج تصنيف يحدث أن يكون له أسماء فئات رقمية. على سبيل المثال، النموذج الذي يتنبأ برمز بريدي رقمي هو نموذج تصنيف وليس نموذج انحدار.

Regularization

التنظيم

أي آلية تقلل الضبط الزائد (overfitting). تشمل الأنواع الشائعة للتسوية ما يلي:

- تنظيم L_1 (L_1 regularization)
- تنظيم L_2 (L_2 regularization)
- تنظيم التسرب (dropout regularization)
- التوقف المبكر (early stopping) (هذه ليست طريقة تنظيم رسمية، ولكن يمكن أن تحد بشكل فعال من الضبط الزائد)

يمكن أيضًا تعريف التنظيم بأنه عقوبة على تعقيد النموذج.

التنظيم على غير المتوقع. عادةً ما تؤدي زيادة التنظيم إلى زيادة خطأ التدريب، وهو أمر محير لأنه، حسنًا، ليس الهدف هو تقليل خطأ التدريب؟

في الواقع لا. الهدف ليس تقليل خطأ التدريب. الهدف هو عمل تنبؤات ممتازة على أمثلة من العالم الحقيقي. من اللافت للنظر، على الرغم من أن زيادة التنظيم تؤدي إلى زيادة خطأ التدريب، إلا أنها تساعد النماذج عادةً على إجراء تنبؤات أفضل في الأمثلة الواقعية.

Regularization rate

معدل التنظيم

رقم يحدد الأهمية النسبية للتنظيم أثناء التدريب. يؤدي رفع معدل التنظيم إلى تقليل الضبط الزائد (overfitting) ولكنه قد يقلل من القدرة التنبؤية للنموذج. على العكس من ذلك، يؤدي تقليل معدل التنظيم أو حذفه إلى زيادة الضبط الزائد.

عادة ما يتم تمثيل معدل التنظيم بالحرف اليوناني (lambda). توضح معادلة الخطأ المبسطة التالية تأثير (lambda):

$$\text{minimize}(\text{loss function} + \lambda(\text{regularization}))$$

حيث يكون التنظيم هو أي آلية تنظيم، بمافي ذلك:

- تنظيم L_1 (L_1 regularization)
- تنظيم L_2 (L_2 regularization)

Reinforcement learning (RL)

التعلم المعزز (RL)

مجموعة من الخوارزميات التي تتعلم السياسة (policy) المثلى، والتي تهدف إلى تعظيم العائد (return) عند التفاعل مع البيئة (environment). على سبيل المثال، المكافأة النهائية لمعظم الألعاب هي الفوز. يمكن أن تصبح أنظمة التعلم المعزز خبيرةفي ممارسة الألعاب المعقدة من خلال تقييم تسلسل حركات اللعبة السابقة التي أدتفي النهاية إلى انتصارات وتسلسلات أدتفي النهاية إلى خسائر.

ReLU

اختصار للوحدة الخطية المصححة (Rectified Linear Unit).

Replay buffer

في خوارزميات تشبه DQN، الذاكرة المستخدمة من قبل الوكيل (agent) لتخزين انتقالات الحالة لاستخدامهافي تجربة إعادة التشغيل (experience replay).

Representation

التمثيل

عملية تعيين البيانات إلى الميزات (features) المفيدة.

Re-ranking

إعادة الترتيب

المرحلة الأخيرة من نظام التوصية (recommendation system)، والتي يمكن خلالها إعادة تصنيف العناصر المسجلة وفقًا لبعض الخوارزمية الأخرى (عادةً، غير خوارزميات التعلم الآلي). تقوم إعادة الترتيب بتقييم قائمة العناصر التي تم إنشاؤها بواسطة مرحلة تسجيل النقاط (scoring)، واتخاذ إجراءات مثل:

- القضاء على العناصر التي قام المستخدم بشرائها بالفعل.
- تعزيز نتيجة العناصر الأكثر حداثة.

Residual Network (ResNet)

الشبكات المتبقية

الشبكة العصبية المتبقية (ResNet) هي شبكة عصبية اصطناعية (ANN). إنه نوع لا يحتوي على بوابة أو مفتوح بوابات من HighwayNet، وهو أول شبكة عصبية عميقة للغاية تعمل على التغذية الأمامية مع مئات الطبقات، أعمق بكثير من الشبكات العصبية السابقة.

Return

العائد

في التعلم المعزز (reinforcement learning)، بالنظر إلى سياسة معينة وحالة معينة، فإن العائد هو مجموع جميع المكافآت (rewards) التي يتوقع الوكيل (agent) الحصول عليها عند اتباع السياسة (policy) من الحالة (state) إلى نهاية الحلقة (episode). يحسب الوكيل الطبيعة المتأخرة للمكافآت المتوقعة من خلال خصم المكافآت وفقًا لتحولات الحالة المطلوبة للحصول على المكافأة.

لذلك، إذا كان عامل الخصم، γ وكان $r_0, \ldots, r_N$ يشير إلى المكافآت حتى نهاية الحلقة، فسيتم حساب العائد على النحو التالي:

$$\text{Return} = r_0 + \gamma r_1 + \gamma^2 r_2 + \ldots + \gamma^{N-1} r_{N-1}$$

Reward

المكافأة

في التعلم المعزز (reinforcement learning)، النتيجة العددية لاتخاذ إجراء (action)في حالة (state) ما، كما تحددها البيئة (environment).

Ridge regularization

تنظيم ريدج

مرادف لتنظيم L_2 (L$_2$ regularization). يتم استخدام مصطلح تنظيم ريدج بشكل متكرر في سياقات الإحصاءات البحتة، بينما يتم استخدام تنظيم L_2 في كثير من الأحيان في التعلم الآلي.

RMSPorp

الانتشار التربيعي لمتوسط الجذر (Root Mean Squared Propagation)، أو RMSProp، هو امتداد للتدرج الاشتقاقي (gradient descent) ونسخة (AdaGrad) من نزول التدرج (gradient) الذي يستخدم متوسطًا متحللًا للتدرجات الجزئية في تكييف حجم الخطوة لكل معلمة.

RNN

اختصار للشبكات العصبية المتكررة (recurrent neural networks).

ROC (receiver operating characteristic) Curve

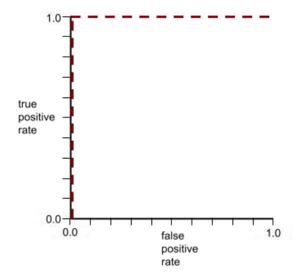
منحنى خصائص تشغيل المستقبل (ROC)

رسم بياني للمعدل الإيجابي الحقيقي (true positive rate) مقابل المعدل الإيجابي الخاطئ (false positive rate) لعتبات التصنيف (classification thresholds) المختلفة في التصنيف الثنائي.

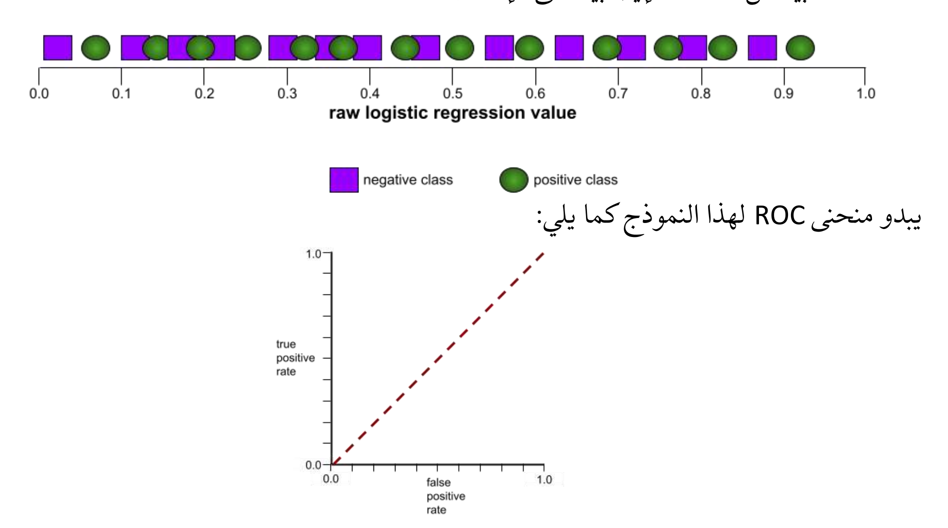
يشير شكل منحنى ROC إلى قدرة نموذج التصنيف الثنائي على فصل الفئات الإيجابية عن الفئات السلبية. افترض، على سبيل المثال، أن نموذج التصنيف الثنائي يفصل تمامًا جميع الفئات السلبية عن جميع الفئات الإيجابية:



يبدو منحنى ROC للنموذج السابق كما يلي:

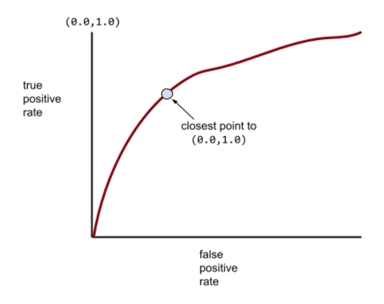


في المقابل، يوضح الرسم التوضيحي التالي قيم الانحدار اللوجستي الخام لنموذج لا يمكنه فصل الفئات السلبية عن الفئات الإيجابية على الإطلاق:



يبدو منحنى ROC لهذا النموذج كما يلي:

في هذه الأثناء، بالعودة إلى العالم الحقيقي، تفصل معظم نماذج التصنيف الثنائي بين الفئات الإيجابية والسلبية إلى حد ما، ولكن عادةً ليس بشكل كامل. لذلك، يقع منحنى ROC النموذجي في مكان ما بين الطرفين:



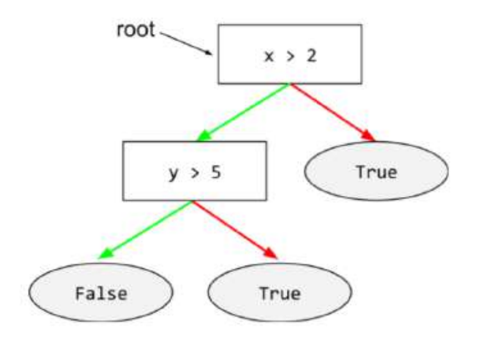
النقطة على منحنى ROC الأقرب إلى (0.0،1.0) تحدد نظريًا عتبة التصنيف المثالية. ومع ذلك، هناك العديد من القضايا الواقعية الأخرى التي تؤثر على اختيار عتبة التصنيف المثالية. على سبيل المثال، ربما تسبب النتائج السلبية الخاطئة ألمًا أكبر بكثير من الإيجابيات الكاذبة.

يلخص المقياس العددي المسمى AUC منحنى ROCفي قيمة نقطة عائمة واحدة.

Root

الجذر

عقدة (node) البداية (الشرط) (condition) الأول)في شجرة القرار (decision tree). حسب الاصطلاح، تضع الرسوم البيانية الجذرفي أعلى شجرة القرار. فمثلا:



Root Mean Squared Error (RMSE)

جذر متوسط الخطأ التربيعي (RMSE)

الجذر التربيعي لمتوسط الخطأ التربيعي (Mean Squared Error).

Rotational invariance

الثبات الدوراني

في مشكلة تصنيف الصور، قدرة الخوارزمية على تصنيف الصور بنجاح حتى عندما يتغير اتجاه الصورة. على سبيل المثال، لا يزال بإمكان الخوارزمية تحديد مضرب التنس سواء كان يشير لأعلى أو جانبيًا أو لأسفل. لاحظ أن الثبات الدوراني ليس مرغوبًا دائمًا؛ على سبيل المثال، لا يجب تصنيف الرقم 9 المقلوب بالرقم 9.

انظر أيضًا ثبات الترجمة (translational invariance) وثبات الحجم (size invariance).

S

S

Scalar

القيمة القياسية

رقم واحد أو سلسلة واحدة يمكن تمثيلها كموتّر (tensor) من الرتبة 0. على سبيل المثال، تُنشئ كل سطور من التعليمات البرمجية التالية قيمة قياسية في TensorFlow:

```
breed = tf.Variable("poodle", tf.string)
temperature = tf.Variable(27, tf.int16)
precision = tf.Variable(0.982375101275, tf.float64)
```

Scalar Multiplication

الضرب القياسي

في الرياضيات، يعد الضرب القياسي أحد العمليات الأساسية التي تحدد مساحة متجه في الجبر الخطي. في السياقات الهندسية الشائعة، الضرب القياسي لمتجه إقليدي حقيقي بعدد حقيقي موجب يضاعف حجم المتجه - دون تغيير اتجاهه.

Scaling

التحجيم

أي تحويل أو أسلوب رياضي يغير نطاق التسمية و / أو قيمة الميزة. بعض أشكال القياس مفيدة جداً للتحولات مثل التسوية (normalization).

تتضمن الأشكال الشائعة من القياس المفيدة في التعلم الآلي ما يلي:

- التحجيم الخطي، والذي يستخدم عادةً مزيجًا من الطرح والقسمة لاستبدال القيمة الأصلية برقم بين 1- و 1+ أو بين 0 و 1.
- التحجيم اللوغاريتمي، الذي يستبدل القيمة الأصلية باللوغاريتم.
- تسوية درجة Z (Z-score normalization)، والتي تحل محل القيمة الأصلية بقيمة فاصلة عائمة تمثل عدد الانحرافات المعيارية عن متوسط تلك الميزة.

scikit-learn

منصة تعلم آلي شهيرة ومفتوحة المصدر.

SciPy

SciPy هي مكتبة Python مجانية ومفتوحة المصدر تُستخدم في الحوسبة العلمية والحوسبة التقنية. يحتوي SciPy على وحدات للتحسين، والجبر الخطي، والتكامل، والاستيفاء، والدوال الخاصة، ومعالجة الإشارات والصور وغيرها من المهام الشائعة في العلوم والهندسة.

Self-attention (also called self-attention layer)

الانتباه الذاتي (طبقة الانتباه الذاتي)

طبقة شبكة عصبية تقوم بتحويل سلسلة من الزخارف (على سبيل المثال، تضمين الرمز المميز (token)) إلى تسلسل آخر من التضمينات (embedding). يتم إنشاء كل تضمين في تسلسل الإخراج من خلال دمج المعلومات من عناصر تسلسل الإدخال من خلال آلية الانتباه.

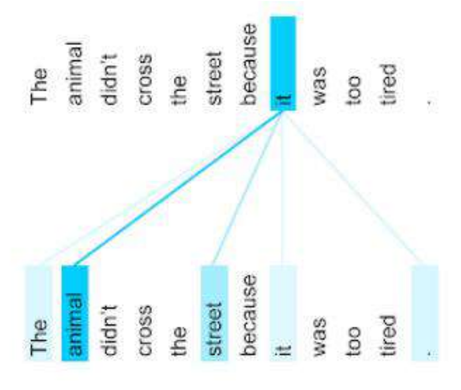
يشير الجزء الذاتي (self) من الانتباه الذاتي إلى التسلسل الذي يهتم بنفسه بدلاً من سياق آخر. الانتباه الذاتي هو أحد الكتل الإنشائية الرئيسية للمحولات (Transformers) ويستخدم مصطلحات البحث في القاموس، مثل "استعلام query" و "مفتاح key" و "قيمة value".

تبدأ طبقة الانتباه الذاتي بسلسلة من تمثيلات المدخلات، واحدة لكل كلمة. يمكن أن يكون تمثيل الإدخال لكلمة تضميناً بسيطاً. لكل كلمة في تسلسل الإدخال، تقوم الشبكة بتسجيل مدى ملاءمة الكلمة لكل عنصر في تسلسل الكلمات بالكامل. تحدد درجات الملاءمة إلى أي مدى يشتمل التمثيل النهائي للكلمة على تمثيلات للكلمات الأخرى.

على سبيل المثال، ضع في اعتبارك الجملة التالية:

```
The animal didn't cross the street because it was too tired.
```

يُظهر الرسم التوضيحي التالي نمط انتباه طبقة الانتباه الذاتي للضمير it، مع تظليل كل سطر يشير إلى مدى مساهمة كل كلمة في التمثيل:



تُبرز طبقة الانتباه الذاتي الكلمات ذات الصلة بـ "it".في هذه الحالة، تعلمت طبقة الانتباه تمييز الكلمات التي قد تشير إليها، مع إعطاء أعلى وزن لـ "animal".

لسلسلة من n رمز مميز (tokens)، يحول الانتباه الذاتي سلسلة من التضمينات n مرات منفصلة، مرة واحدةفي كل موضعفي التسلسل.

راجع أيضًا الانتباه (attention) والانتباه الذاتي متعدد الرؤوس (multi-head self-attention).

Self-supervised learning

التعلم تحت الإشراف الذاتي

مجموعة من التقنيات لتحويل مشكلة التعلم الآلي غير الخاضعة للإشراف (unsupervised machine learning) إلى مشكلة تعلم آلي خاضعة للإشراف (supervised machine learning) من خلال إنشاء تسميات بديلة من أمثلة غير مسماة.

تستخدم بعض النماذج المستندة إلى المحولات (Transformer) مثل (BERT)التعلم تحت الإشراف الذاتي.

التدريب تحت الإشراف الذاتي هو نهج تعلم شبه خاضع للإشراف (semi-supervised learning).

Self-training

التدريب الذاتي

نوع من التعلم تحت الإشراف الذاتي (self-supervised learning) يكون مفيداً بشكل خاص عندما تكون جميع الشروط التالية صحيحة:

- نسبة الأمثلة غير المسماة (unlabeled examples) إلى الأمثلة المسماة (labeled examples)في مجموعة البيانات عالية.
- هذه مشكلة تصنيف (classification).

يعمل التدريب الذاتي من خلال التكرار على الخطوتين التاليتين حتى يتوقف النموذج عن التحسن:

- استخدم التعلم الآلي الخاضع للإشراف (supervised machine learning) لتدريب نموذج على الأمثلة المسماة.

- استخدم النموذج الذي تم إنشاؤهفي الخطوة 1 لإنشاء تنبؤات (تسميات) على الأمثلة غير المسماة، ونقل تلك التي توجد فيها ثقة عاليةفي الأمثلة المسماة مع التسمية المتوقعة.

لاحظ أن كل تكرار للخطوة 2 يضيف المزيد من الأمثلة المسماة للخطوة 1 للتدريب عليها.

Segmentation

التقطيع

إنها عملية تقسيم مجموعة البيانات إلى مجموعات متميزة متعددة. يتم هذا الفصل بحيث يتشابه أعضاء المجموعة نفسها مع بعضهم البعض ويختلفون عن أعضاء المجموعات الأخرى.

Semantic Segmentation

التجزئة الدلالية

التجزئة الدلالية هي خوارزمية تعلم عميق تربط تسمية أو فئة بكل بكسلفي الصورة. يتم استخدامه للتعرف على مجموعة من وحدات البكسل التي تشكل فئات متميزة.

Semi-supervised learning

التعلم شبه الخاضع للإشراف

تدريب نموذج على البيانات حيث تحتوي بعض الأمثلة التدريبية على تسميات (labels) ولكن البعض الآخر لا يحتوي على تسميات. تتمثل إحدى تقنيات التعلم شبه الخاضع للإشرافي استنتاج تسميات الأمثلة غير المسماة (unlabeled examples)، ثم التدريب على التسميات المستنتجة لإنشاء نموذج جديد. يمكن أن يكون التعلم شبه الخاضع للإشراف مفيداً إذا كان الحصول على التسميات مكلف ولكن الأمثلة غير المصنفة وفيرة.

التدريب الذاتي (Self-training) هو أحد أساليب التعلم شبه الخاضع للإشراف.

Sentiment analysis

تحليل المشاعر

استخدام الخوارزميات الإحصائية أو خوارزميات التعلم الآلي لتحديد الموقف العام للمجموعة (إيجابيًا أو سلبيًا) تجاه خدمة أو منتج أو مؤسسة أو موضوع. على سبيل المثال، باستخدام فهم اللغة الطبيعية (natural language understanding)، يمكن لخوارزمية إجراء تحليل المشاعر على التعليقات النصية من دورة جامعية لتحديد الدرجة التي يحبها الطلاب بشكل عام أو يكرهون هذه الدورة.

Sensitivity

الحساسية

يتم تحديد حساسية المصنف من خلال معدله الإيجابي الحقيقي (TPR):

```
sensitivity = TPR = TP/(TP+FN).
```

Seq2Seq Model

نموذج Seq2Seq

نموذج Seq2Seq هو نموذج يأخذ سلسلة من العناصر (كلمات، حروف، سلاسل زمنية، إلخ) ويخرج تسلسل آخر من العناصر.في حالة الترجمة الآلية العصبية، يكون الإدخال عبارة عن سلسلة من الكلمات، والمخرج هو سلسلة الكلمات المترجمة.

Sequence model

نموذج التسلسل

نموذج له مدخلاته تبعية متسلسلة. على سبيل المثال، التنبؤ بالفيديو التالي الذي تمت مشاهدته من خلال سلسلة من مقاطع الفيديو التي تمت مشاهدتها مسبقًا.

Sequence-to-sequence task

مهمة التسلسل إلى التسلسل

مهمة تقوم بتحويل تسلسل إدخال من الرموز المميزة (tokens) إلى تسلسل إخراج من الرموز المميزة. على سبيل المثال، هناك نوعان شائعان من مهام التسلسل إلى التسلسل وهما:

- المترجمون (Translators):
  - نموذج تسلسل إدخال: " أنا أحبك ".
  - عينة تسلسل الإخراج: "Je t'aime".
- إجابة السؤال (Question answering):
  - نموذج تسلسل إدخال: "هل أحتاج سيارتي في مدينة نيويورك؟"
  - نموذج تسلسل الإخراج: "لا. الرجاء إبقاء سيارتك في المنزل".

Shape (Tensor)

عدد العناصر في كل بُعد (dimension) للموتر (tensor). يتم تمثيل الشكل كقائمة من الأعداد الصحيحة. على سبيل المثال، الموتر ثنائي الأبعاد التالي له شكل [3،4]:

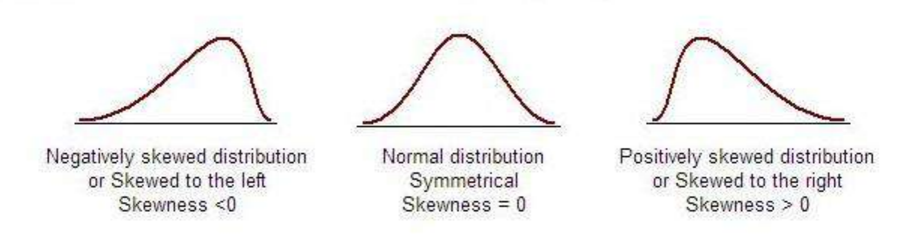
```
[[5, 7, 6, 4],
 [2, 9, 4, 8],
 [3, 6, 5, 1]]
```

يستخدم TensorFlow تنسيق الصف الرئيسي (نمط C) لتمثيل ترتيب الأبعاد، وهذا هو السبب في أن الشكل في TensorFlow هو [3،4] بدلاً من [4،3]. بمعنى آخر،في موتر TensorFlow ثنائي الأبعاد، يكون الشكل هو [عدد الصفوف، عدد الأعمدة].

Skewness

الانحراف

الانحراف مقياس للتماثل (symmetry). يكون التوزيع، أو مجموعة البيانات، متماثلًا إذا كان يبدو متماثلًا على يسار ويمين النقطة المركزية.



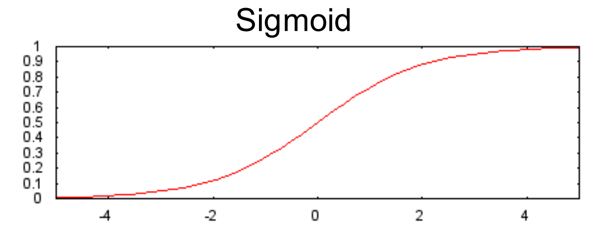
Shrinkage

الانكماش

معلمة فائقة (hyperparameter)في تعزيز التدرج (gradient boosting) الذي يتحكم في الضبط الزائد (overfitting). الانكماش في تعزيز التدرج يماثل معدل التعلم (learning rate) في التدرج الاشتقاقي (gradient descent). الانكماش هو قيمة عشرية بين 0.0 و 1.0. تقلل قيمة الانكماش الأقل من الضبط الزائد أكثر من قيمة الانكماش الأكبر.

Sigmoid

دالة رياضية "تسحق" قيمة إدخال في نطاق مقيد، عادة من 0 إلى 1 أو 1– إلى 1+. أي أنه يمكنك تمرير أي رقم (اثنان، مليون، سالب مليار، أيًا كان) إلى sigmoid وسيظل الناتج في النطاق المقيد. تبدو قطعة sigmoid دالة كما يلي:



لدالة sigmoid استخدامات عديدةفي التعلم الآلي، بمافي ذلك:

- تحويل الناتج الخام للانحدار اللوجستي (logistic regression) أو الانحدار متعدد الحدود (multinomial regression) إلى احتمال.
- العمل كدالة تنشيط (activation function)في بعض الشبكات العصبية.

تحتوي دالة sigmoid على رقم الإدخال x على الصيغة التالية:

$$sigmoid(x) = \frac{1}{1 + e^{-x}}$$

في التعلم الآلي، تعتبر x بشكل عام مجموع مرجح (weighted sum).

Skip-Gram

تعد Skip-gram إحدى تقنيات التعلم غير الخاضعة للإشراف (unsupervised learning) المستخدمة للعثور على أكثر الكلمات ذات الصلة بكلمة معينة. يستخدم Skip-gram للتنبؤ بكلمة سياق لكلمة مستهدفة معينة.

SKLearn

مرادف لـ (scikit-learn).

Sliding-Window

النافذة المنزلقة

تشير النافذة المنزلقةفي الرؤية الحاسوبية (computer vision) إلى منطقة مستطيلة ذات عرض وارتفاع محددين تتحرك فوق الصورة. تلعب النوافذ المنزلقة دورًا مهمًافي تصنيف الكائنات، لأنها تتيح لنا تحديد "مكان" الكائن بالضبطفي الصورة.

Similarity measure

مقياس التشابه

في خوارزميات التكتل (clustering)، يتم استخدام المقياس لتحديد مدى تشابه أي مثالين.

Size invariance

ثبات الحجم

في مشكلة تصنيف الصور، قدرة الخوارزمية على تصنيف الصور بنجاح حتى عندما يتغير حجم الصورة. على سبيل المثال، لا يزال بإمكان الخوارزمية تحديد قطة سواء كانت تستهلك 2 ميغا بكسل أو 200 ألف بكسل. لاحظ أنه حتى أفضل خوارزميات تصنيف الصور لا تزال لها حدود

عملية على ثبات الحجم. على سبيل المثال، من غير المحتمل أن تقوم خوارزمية (أو بشرية) بتصنيف صورة قطة تستهلك 20 بكسل فقط بشكل صحيح.

انظر أيضًا ثبات الترجمة (translational invariance) وثبات الدوران (rotational invariance).

Sketching

التخطيط

في التعلم الآلي غير الخاضع للإشراف (unsupervised machine learning)، فئة من الخوارزميات تقوم بإجراء تحليل أولي للتشابه على الأمثلة. تستخدم خوارزميات التخطيط دالة تجزئة حساسة للمنطقة المحلية (locality-sensitive hash function) لتحديد النقاط التي من المحتمل أن تكون متشابهة، ثم تجميعهافي مجموعات.

يقلل التخطيط من الحساب المطلوب لحسابات التشابه على مجموعات البيانات الكبيرة. بدلاً من حساب التشابه لكل زوج من الأمثلةفي مجموعة البيانات، نحسب التشابه فقط لكل زوج من النقاط داخل كل مجموعة.

SMOTE

وهي عبارة عن أسلوب لأخذ العينات من الأقليات التركيبة (Synthetic Minority Over-Sampling) وهو نهج لبناء المصنفات من مجموعات البيانات غير المتوازنة (imbalanced datasets) الموصوفة. الفكرة الكامنة وراء هذه التقنية هي أن الإفراطفي أخذ عينات من فئة الأقلية (minority) (غير الطبيعية) وأخذ عينات أقل من فئة الأغلبية (majority) (العادية) يمكن أن تحقق أداء مصنفاً أفضل (في مساحة ROC) من أخذ عينات أقل من فئة الأغلبية فقط. هذا هو نهج أخذ العينات المفرط (oversampling) حيث يتم أخذ عينات من فئة الأقلية بشكل مفرط عن طريق إنشاء أمثلة "تركيبة synthetic " بدلاً من الإفراطفي أخذ العينات مع الاستبدال.

Soft Margin

الهامش الناعم

يسمح هامش الناعمفي خوارزمية الة المتجهات الداعمة (SVM) بحدوث بعض سوء التصنيف عن طريق تخفيف القيود الصعبة لـ SVM. يتم تنفيذ SVM للهامش الناعم بمساعدة معامل التنظيم (C). معامل التنظيم (C): يخبرنا عن مقدار سوء التصنيف الذي نريد تجنبه.

Softmax

دالة تحدد الاحتمالات لكل فئة ممكنةفي نموذج تصنيف متعدد الفئات (multi-class classification model). مجموع الاحتمالات يصل بالضبط إلى 1.0. على سبيل المثال، يوضح الجدول التالي كيف تقوم softmax بتوزيع الاحتمالات المختلفة:

Image is a...	Probability
dog	.85
cat	.13
horse	.02

يسمى Softmax أيضًا بـ (full softmax).

Sparse feature

ميزة متناثرة

عنصر يغلب أن تكون قيمه صفرًا أو فارغًا. على سبيل المثال، الميزة التي تحتوي على قيمة واحدة و0 مليون قيمة متناثرة.في المقابل، تحتوي الميزة الكثيفة (dense feature) على قيمفي الغالب ليست صفرية أو فارغة.

في التعلم الآلي، هناك عدد مذهل من الميزات هي ميزات متناثرة. الميزات الفئوية (Categorical features) عادةً ما تكون ميزات متناثرة. على سبيل المثال، من بين 300 نوع من الأشجار المحتملةفي الغابة، قد يحدد مثال واحد شجرة القيقب فقط. أو، من بين ملايين مقاطع الفيديو المحتملةفي مكتبة الفيديو، قد يحدد مثال واحد "الدار البيضاء" فقط.

في النموذج، عادةً ما تمثل ميزات متناثرة بترميز واحد ساخن (one-hot encoding). إذا كان الترميز واحدًا ساخنًا كبيرًا، فيمكنك وضع طبقة التضمين (embedding layer) أعلى الترميز واحد ساخن لزيادة الكفاءة.

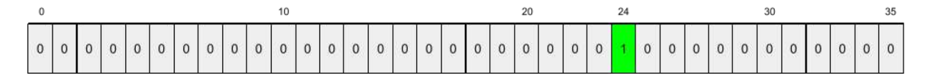
Sparse representation

التمثيل المتناثر

تخزين موضع (مواضع) العناصر غير الصفرية فقطفي ميزة متناثرة (sparse feature).

على سبيل المثال، افترض أن الميزة الفئوية المسماة الأنواع تحدد 36 نوعًا من الأشجارفي غابة معينة. افترض كذلك أن كل مثال يحدد نوعًا واحدًا فقط.

يمكنك استخدام متجه واحد ساخن لتمثيل أنواع الأشجارفي كل مثال. سيحتوي متجه واحد ساخن على 1 واحد (لتمثيل أنواع الأشجار المعينةفي هذا المثال) و 35 0 ثانية (لتمثيل 35 نوعًا من الأشجار غير الموجودةفي هذا المثال). لذلك، قد يبدو التمثيل الواحد الساخن للقيقب كما يلي:



بدلاً من ذلك، فإن التمثيل المتناثر سيحدد ببساطة موقف الأنواع المعينة. إذا كان القيقبفي الموضع 24، فإن التمثيل المتناثر للقيقب سيكون ببساطة:

```
24
```

لاحظ أن التمثيل المتناثر أكثر إحكاما من التمثيل الساخن.

Sparse vector

متجه متناثر

متجه معظم قيمه أصفار. راجع أيضًا الميزة المتناثرة (sparse feature) والتناثر (sparsity).

Sparsity

التناثر

عدد العناصر التي تم ضبطها على صفر (أو لا شيء)في متجه أو مصفوفة مقسومًا على العدد الإجمالي للمدخلاتفي هذا المتجه أو المصفوفة. على سبيل المثال، ضعفي اعتبارك مصفوفة من 100 عنصر تحتوي فيها 98 خلية على صفر. حساب التناثر كالتالي:

$$\text{sparsity} = \frac{98}{100} = 0.98$$

يشير تناثر الميزات (Feature sparsity) إلى تباين متجه الميزة؛ يشير تناثر النموذج (model sparsity) إلى تباين أوزان النموذج.

Spatial pooling

انظر التجميع (pooling).

Specificity

الخصوصية

تشير الخصوصية إلى المعدل السلبي الحقيقي للمصنف:

```
specificity = 1 - TP / (TP + FP) = 1 - FPR
```

Spectogram

المخطط الطيفي

يعد المخطط الطيفي "لقطة" موجزة لموجة صوتية ولأنها صورة، فهي مناسبة تمامًا للإدخال إلى البنى القائمة على CNN التي تم تطويرها للتعامل مع الصور. يتم إنشاء الطيف من الإشارات الصوتية باستخدام تحويلات فورييه.

splitter

المقسم

أثناء تدريب شجرة القرار (decision tree)، يكون الروتين (والخوارزمية) مسؤولين عن إيجاد أفضل حالة (condition)في كل عقدة (node).

Squared hinge loss

خطأ هينج التربيعية

يؤدي خطأ هينج التربيعية إلى معاقبة القيم المتطرفة (outliers) بشكل أكثر شدة من خطأ هينج المنتظم.

Squared loss

الخطأ التربيعي

مرادف لخطأ L_2 (L_2 loss).

Staged training

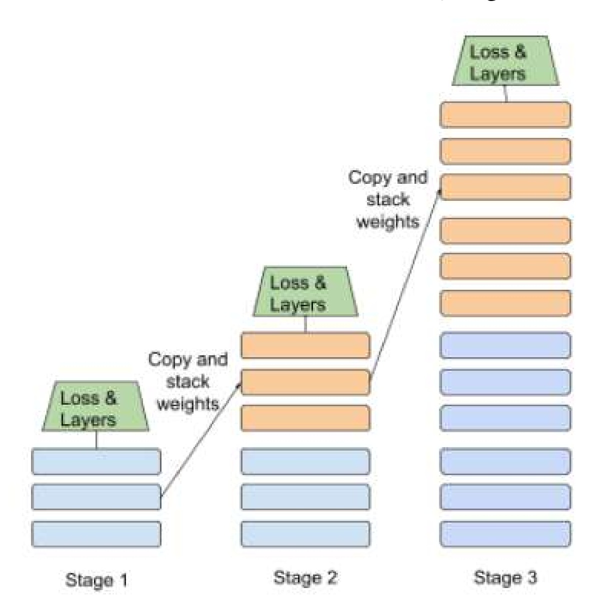
التدريب على مراحل

تكتيك تدريب نموذجفي سلسلة من المراحل المنفصلة. يمكن أن يكون الهدف إما تسريع عملية التدريب، أو تحقيق جودة أفضل للنموذج.

فيما يلي توضيح التدريب على مراحل:

- المرحلة الاولى تحتوي على 3 طبقات مخفية، المرحلة الثانية تحتوي على 6 طبقات مخفية، والمرحلة الثالثة تحتوي على 12 طبقة مخفية.

- تبدأ المرحلة الثانية التدريب بالأوزان التي تم تعلمهافي الطبقات المخفية الثلاثة من المرحلة الأولى. تبدأ المرحلة الثالثة بالتدريب بالأوزان التي تم تعلمهافي الطبقات المخفية الستة من المرحلة الثانية.



انظر أيضا (pipelining).

Standard Deviation

الانحراف المعياري

يشير الانحراف المعياري إلى مدى تشتت البيانات. إنه الجذر التربيعي لتباين البيانات الأساسية. يتم حساب الانحراف المعياري للمجموعة.

Standardization

التوحيد القياسي

التوحيد القياسي (أو تسوية درجة Z (Z-score normalization)) هو العملية التي يتم فيها إعادة قياس الميزات بحيث يكون لها خصائص التوزيع الطبيعي القياسي مع $\mu = 0$ و$\sigma = 1$، حيث μ هي المتوسط (mean) وσ هي الانحراف المعياري عن المتوسط. يتم حساب الدرجات القياسية (تسمى أيضًا درجات z) للعينات على النحو التالي:

$$z = \frac{x - \mu}{\sigma}$$

State

الحالة

في التعلم المعزز (reinforcement learning)، قيم المعلمات التي تصف التكوين الحالي للبيئة، والتي يستخدمها الوكيل (agent) لاختيار إجراء (action).

State-action value function

دالة قيم الحالة-الاجراء

مرادف لدالة Q (Q-function).

Static

ثابت

شيء يتم عمله مرة واحدة وليس بشكل مستمر. المصطلحان ثابت (static) واوفلاين (offline) هما مترادفان. فيما يلي الاستخدامات الشائعة للثابت والاوفلاين في التعلم الآلي:

- النموذج الثابت (static model) (أو النموذج الاوفلاين (offline model)) هو نموذج تم تدريبه مرة واحدة ثم استخدامه لفترة من الوقت.
- التدريب الثابت (static training) (أو التدريب الاوفلاين (offline training)) هو عملية تدريب نموذج ثابت.
- الاستدلال الثابت (static inference) (أو الاستدلال الاوفلاين (offline inference)) هو عملية يولد فيها النموذج مجموعة من التنبؤات في وقت واحد.

على النقيض من الديناميكي (dynamic).

Static inference

الاستدلال الثابت

مرادف للاستدلال الاوفلاين (offline inference).

Stationarity

الثبات

ميزة لا تتغير قيمها عبر بُعد واحد أو أكثر، عادةً ما يكون الوقت. على سبيل المثال، تُظهر الميزة التي تبدو قيمها متشابهة في 2020 و 2022 الثبات.

في العالم الحقيقي، هناك عدد قليل جداً من الميزات التي تظهر ثابتة. حتى الميزات المرادفة للاستقرار (مثل مستوى سطح البحر) تتغير بمرور الوقت.

على النقيض من (nonstationarity).

Stemming

التجذيع

التجذيع هي تقنية معالجة لغة طبيعية (NLP) تعمل على تقليل انعطاف الكلمات إلى أشكالها الجذرية، وبالتالي المساعدةفي المعالجة المسبقة للنص والكلمات والمستندات لتطبيع النص.

Step

خطوة

تمريرة أمامية وخلفية دفعة (batch) واحدة.

انظر الانتشار الخلفي (backpropagation) لمزيد من المعلومات حول التمريرة الأمامية (forward pass) والتمريرة الخلفة (backward pass).

Step Size

حجم الخطوة

مرادف لمعدل التعلم (learning rate).

Stochastic gradient descent (SGD)

التدرج الاشتقاقي العشوائي (SGD)

خوارزمية التدرج الاشتقاقي (gradient descent) التي يكون فيها حجم الدُفعة (batch size) واحداً. بمعنى آخر، تدرب SGD على مثال واحد يتم اختياره بشكل موحد عشوائيًا من مجموعة تدريب (training set).

Stride

الخطوة

في عملية تلافيفية (convolutional operation) أو تجميع (pooling)، دلتافي كل بُعد من أبعاد السلسلة التالية من شرائح الإدخال. على سبيل المثال، يوضح الرسم المتحرك التالي خطوة (1،1) أثناء عملية تلافيفية. لذلك، تبدأ شريحة الإدخال التاليةفي موضع واحد على يمين شريحة

الإدخال السابقة. عندما تصل العملية إلى الحافة اليمنى، تكون الشريحة التالية على طول الطريق إلى اليسار ولكن موضعًا واحدًا لأسفل.

128	97	53	201	198
35	22	25	200	195
37	24	28	197	182
33	28	92	195	179
31	40	100	192	177

181		

يوضح المثال السابق خطوة ثنائية الأبعاد. إذا كانت مصفوفة الإدخال ثلاثية الأبعاد، فإن الخطوة ستكون أيضًا ثلاثية الأبعاد.

Structural risk minimization (SRM)

تقليل المخاطر الهيكلية (SRM)

خوارزمية توازن بين هدفين:

- الرغبة في بناء النموذج الأكثر تنبوًا (على سبيل المثال، أقل خطأ).
- الرغبة في إبقاء النموذج بسيطًا قدر الإمكان (على سبيل المثال، تنظيم قوي).

على سبيل المثال، الدالة التي تقلل من الخطأ + التنظيم في مجموعة التدريب هي خوارزمية لتقليل المخاطر الهيكلية.

على النقيض من تقليل المخاطر التجريبية (empirical risk minimization).

Style Transfer

نقل النمط

يشير أسلوب نقل النمط إلى فئة من خوارزميات البرامج التي تتعامل مع الصور الرقمية أو مقاطع الفيديو من أجل اعتماد المظهر أو النمط المرئي لصورة أخرى. تتميز خوارزميات نقل النمط باستخدامها للشبكات العصبية العميقة من أجل تحويل الصورة.

Subsampling

أخذ عينات فرعية

انظر التجميع (pooling).

Supervised machine learning

التعلم الآلي الخاضع للإشراف

تدريب نموذج (model) من الميزات (features) والتسميات (labels) المقابلة لها. يشبه التعلم الآلي الخاضع للإشراف تعلم موضوع ما من خلال دراسة مجموعة من الأسئلة والإجابات المقابلة لها. بعد إتقان التعيين بين الأسئلة والأجوبة، يمكن للطالب بعد ذلك تقديم إجابات لأسئلة جديدة (لم يسبق رؤيتها من قبل) حول نفس الموضوع.

قارن مع التعلم الآلي غير الخاضع للإشراف (unsupervised machine learning).

Support Vector Machine (SVM)

الة المتجهات الداعمة

في التعلم الآلي، الة المتجهات الداعمة هي نماذج تعلم خاضعة للإشراف مع خوارزميات التعلم المرتبطة التي تحلل البيانات من أجل التصنيف وتحليل الانحدار.

Synthetic feature

ميزة تركيبية

ميزة (feature) غير موجودة بين ميزات الإدخال، ولكن تم تجميعها من واحدة أو أكثر منها. تتضمن طرق إنشاء الميزات التركيبية ما يلي:

- تجميع (Bucketing) ميزة مستمرةفي صناديق النطاق.
- خلق ميزة التقاطع (feature cross).
- ضرب (أو قسمة) قيمة ميزة واحدة على قيمة (قيم) ميزة أخرى أو بمفردها. على سبيل المثال، إذاكانت a و b من سمات الإدخال، فإن ما يلي هو أمثلة على الميزات التركيبية:

    - `ab`
    - a^2

- تطبيق دالة على قيمة الميزة. على سبيل المثال، إذاكانت c ميزة إدخال، فإن ما يلي هو أمثلة على الميزات التركيبية:

    - `sin(c)`
    - `ln(c)`

لا تعتبر الميزات التي تم إنشاؤها عن طريق التسوية (normalizing) أو القياس (scaling) بمفردها ميزات تركيبية.

T

T

Tabular Q-learning

تعلم Q الجدولي

في التعلم المعزز (reinforcement learning)، تنفيذ تعلم Q (Q-learning) باستخدام جدول لتخزين دوال Q (Q-functions) لكل مجموعة من الحالات (states) والاجراء (action).

Tanh

يشار أيضًا إلى دالة تنشيط الظل الزائدي (hyperbolic tangent activation) ببساطة باسم دالة Tanh (أيضًا "tanh" و "TanH"). إنها تشبه إلى حد بعيد دالة (sigmoid) ولها نفس الشكل S. تأخذ الدالة أي قيمة حقيقية كقيم مدخلات ومخرجات في النطاق من 1 ـ إلى 1.

Target

مرادف للتسمية (label).

Target network

الشبكة المستهدفة

في تعلم Q العميق (Deep Q-Learning)، شبكة عصبية هي تقريب مستقر للشبكة العصبية الرئيسية، حيث تنفذ الشبكة العصبية الرئيسية إما دالة Q (Q-function) أو سياسة (policy). بعد ذلك، يمكنك تدريب الشبكة الرئيسية على قيم Q التي تنبأت بها الشبكة المستهدفة. لذلك، فإنك تمنع حلقة التغذية الراجعة التي تحدث عندما تتدرب الشبكة الرئيسية على قيم Q التي تنبأ بها بنفسها. من خلال تجنب هذه التغذية الراجعة، يزيد ثبات التدريب.

Temporal data

البيانات الزمنية

البيانات المسجلة في نقاط زمنية مختلفة. على سبيل المثال، ستكون مبيعات المعاطف الشتوية المسجلة لكل يوم من أيام السنة بيانات زمنية.

Tensor

الموتر

هيكل البيانات الأساسي في برامج (TensorFlow). الموترات هي هياكل بيانات ذات أبعاد N (حيث يمكن أن تكون كبيرة جداً)، وهي في الغالب مقاييس أو متجهات أو مصفوفات. يمكن أن تحتوي عناصر الموترات (Tensor) على عدد صحيح، أو نقطة عائمة، أو قيم سلسلة.

TensorBoard

لوحة المعلومات التي تعرض الملخصات المحفوظة أثناء تنفيذ برنامج أو أكثر من برامج TensorFlow.

TensorFlow

TensorFlow هي مكتبة برامج مجانية ومفتوحة المصدر لتعلم الآلة والذكاء الاصطناعي. يمكن استخدامه عبر مجموعة من المهام ولكن يركز بشكل خاص على تدريب واستدلال الشبكات العصبية العميقة.

TensorFlow Playground

برنامج يصور كيف تؤثر المعلمات الفائقة (hyperparameters) المختلفة على تدريب النموذج (الشبكة العصبية بشكل أساسي). انتقل إلى http://playground.tensorflow.org/ لتجربة TensorFlow Playground.

TensorFlow Serving

منصة لنشر النماذج المدربة في الإنتاج.

Tensor Processing Unit (TPU)

دائرة متكاملة خاصة بالتطبيقات (ASIC) تعمل على تحسين أداء أعباء عمل التعلم الآلي. يتم نشر ASICs هذه كرقاقات TPU متعددة على جهاز TPU.

Tensor shape

عدد العناصر التي يحتوي عليها الموتر (Tensor) بأبعاد مختلفة. على سبيل المثال، الموتر [5، 10] له شكل 5 في بُعد واحد و10 في بُعد آخر.

Tensor size

العدد الإجمالي للكميات التي يحتوي عليها الموتر (Tensor). على سبيل المثال، حجم موتر [5 ، 10] هو 50.

Termination condition

شرط الإنهاء

في التعلم المعزز (reinforcement learning)، الشروط التي تحدد متى تنتهي الحلقة (episode)، مثل عندما يصل العامل إلى حالة معينة أو يتجاوز عدد عتبة انتقالات الحالة. على سبيل المثال،في tic-tac-toe (المعروف أيضًا باسم noughts والصلبان)، تنتهي الحلقة إما عندما يحدد اللاعب ثلاث مسافات متتالية أو عندما يتم تحديد جميع المسافات.

Test

الاختبار

في شجرة القرار (decision tree)، اسم آخر للشرط (condition).

Test loss

خطأ الاختبار

مقياس (metric) يمثل خطأ (loss) النموذج مقابل مجموعة الاختبار (test set). عند إنشاء نموذج، تحاول عادةً تقليل خطأ الاختبار إلى الحد الأدنى. هذا لأن الخطأ المنخفض في الاختبار هي إشارة جودة أقوى من خطأ التدريب (training loss) المنخفض أو خطأ التحقق من الصحة (validation loss) المنخفض.

تشير الفجوة الكبيرة بين خطأ الاختبار وخطأ التدريب أو خطأ التحقق أحيانًا إلى أنك بحاجة إلى زيادة معدل التنظيم (regularization rate).

Test set

مجموعة الاختبار

مجموعة فرعية من مجموعة البيانات محجوزة لاختبار نموذج مدرب.

تقليديا، تقسم الأمثلةفي مجموعة البيانات إلى المجموعات الفرعية الثلاثة المتميزة التالية:

- مجموعة التدريب (training set).
- مجموعة التحقق من الصحة (validation set).
- مجموعة الاختبار (test set).

يجب أن ينتمي كل مثالفي مجموعة البيانات إلى واحدة فقط من المجموعات الفرعية السابقة. على سبيل المثال، يجب ألا ينتمي مثال واحد إلى كل من مجموعة التدريب ومجموعة الاختبار.

ترتبط مجموعة التدريب ومجموعة التحقق ارتباطًا وثيقًا بتدريب النموذج. نظرًا لأن مجموعة الاختبار مرتبطة بشكل غير مباشر فقط بالتدريب، فإن خطأ الاختبار (test loss) هي مقياس أقل تحيزًا وأعلى جودة من خطأ التدريب (training loss) أو خطأ التحقق من الصحة (validation loss).

Text Mining

التنقيب عن النص

التنقيب عن النص هو عملية استخلاص معلومات عالية الجودة من النص. وهي تنطوي على "اكتشاف الكمبيوتر لمعلومات جديدة غير معروفة من قبل، عن طريق الاستخراج التلقائي للمعلومات من مصادر مكتوبة مختلفة.

TF-IDF

تردد الكلمة- تردد المُستند العكسي

هو إحصاء رقمي يهدف إلى عكس مدى أهمية كلمة ما لوثيقةفي مجموعة نصية (corpus). غالبًا ما يتم استخدامه كعامل ترجيحفي عمليات البحث عن استرجاع المعلومات (information retrieval)، والتنقيب عن النص (text mining)، ونمذجة المستخدم (user modeling).

tf.keras

تم دمج تطبيق Kerasفي TensorFlow.

Threshold (for decision trees)

في حالة محاذاة المحور (axis-aligned condition)، القيمة التي تتم مقارنة الميزة (feature) بها. على سبيل المثال، 75 هي قيمة الحدفي الحالة التالية:

```
grade >= 75
```

Time series analysis

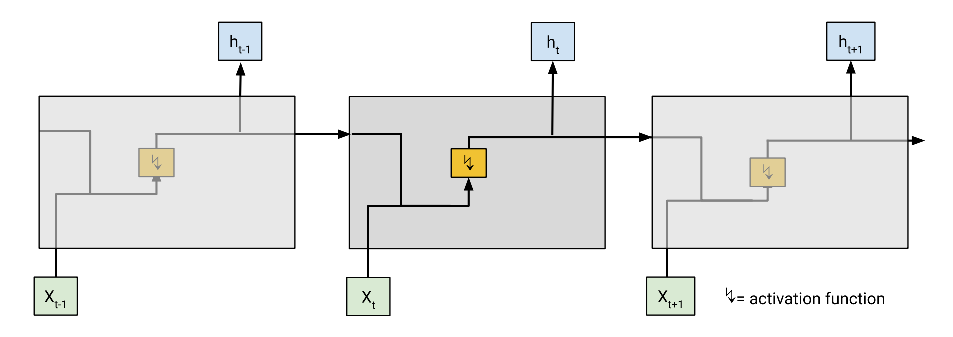
تحليل السلاسل الزمنية

حقل فرعي من التعلم الآلي والإحصاءات التي تحلل البيانات الزمنية (temporal data). تتطلب العديد من أنواع مشكلات التعلم الآلي تحليل السلاسل الزمنية، بمافي ذلك التصنيف والتجميع والتنبؤ واكتشاف الانحراف. على سبيل المثال، يمكنك استخدام تحليل السلاسل الزمنية للتنبؤ بالمبيعات المستقبلية للمعاطف الشتوية حسب الشهر بناءً على بيانات المبيعات التاريخية.

Timestep

خطوة زمنية

خلية واحدة "unrolled" داخل شبكة عصبية متكررة (recurrent neural network). على سبيل المثال، يوضح الشكل التالي ثلاث خطوات زمنية (معنونة بالرموز الفرعية t-1 و t و t + 1):



Token

وحدة لفظية (الرمز المميز)

في نموذج لغوي (language model)، الوحدة الذرية التي يتدرب عليها النموذج ويقوم بالتنبؤات. عادةً ما يكون وحدة لفظية واحداً مما يلي:

- كلمة: على سبيل المثال، تتكون عبارة "dogs like cats" من ثلاث كلمات مميزة: "dogs" و " like " و "cats".
- حرف: على سبيل المثال، تتكون عبارة "bike fish" من تسع وحدات لفظية. (لاحظ أن المساحة الفارغة تُحسب كأحد الوحدات اللفظية)
- الكلمات الفرعية: حيث يمكن أن تكون كلمة واحدة وحدة لفظية واحدةً أو وحدات لفظية متعددة. تتكون الكلمة الفرعية من كلمة جذر أو بادئة أو لاحقة. على سبيل المثال، قد يعرض نموذج اللغة الذي يستخدم كلمات فرعية كرموز لكلمة "dogs" كوحدتين لفظيتين (جذر الكلمة "dog" ولاحقة الجمع "s"). قد يعرض نموذج اللغة نفسه الكلمة المفردة "taller" ككلمتين فرعيتين (الكلمة الجذر "tall" واللاحقة "er").

في المجالات خارج نماذج اللغة، يمكن أن تمثل الوحدات اللفظية أنواعًا أخرى من كل الوحدات. على سبيل المثال،في الرؤية الحاسوبية، قد تكون الوحدة اللفظية مجموعة فرعية من الصورة.

Tokenization

التقسيم الى وحدات لغوية

يتم استخدام Tokenizationفي معالجة اللغة الطبيعية (NLP) لتقسيم الفقرات والجمل إلى وحدات أصغر يمكن تعيين معنى لها بسهولة أكبر.

Torch

Torch هي مكتبة مفتوحة المصدر للتعلم الآلي ، تعتمد على لغة برمجة Lua. يوفر مجموعة واسعة من الخوارزميات للتعلم العميق.

Tower

البرج

أحد مكونات الشبكة العصبية العميقة (deep neural network) التي هي نفسها شبكة عصبية عميقة بدون طبقة مخرجات. عادةً ما يقرأ كل برج من مصدر بيانات مستقل. الأبراج مستقلة حتى يتم دمج مخرجاتهافي طبقة نهائية.

TPU

اختصار لوحدة معالجة الموتر (Tensor Processing Unit).

Training

التدريب

عملية تحديد المعلمات (parameters) المثالية (الأوزان (weights) والتحيزات (biases)) التي تتألف منها النموذج (model). أثناء التدريب، يقرأ النظامفي الأمثلة (examples) ويضبط المعلمات تدريجيًا. يستخدم التدريب كل مثالفي أي مكان من عدة مرات إلى مليارات المرات.

Training loss

خطأ التدريب

مقياس (metric) يمثل خطأ (loss) النموذج أثناء تكرار تدريب معين. على سبيل المثال، افترض أن دالة الخطأ هي متوسط الخطأ التربيعي (Mean Squared Error). ربما تكون خطأ التدريب (متوسط الخطأ التربيعي) للتكرار العاشر 2.2، وخطأ التدريب للتكرار 100 هي 1.9.

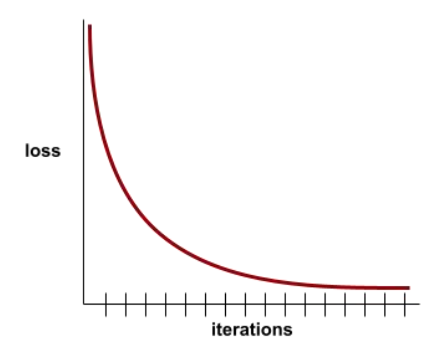
منحنى الخطأ (loss curve) يرسم خطأ التدريب مقابل عدد التكرارات. يوفر منحنى الخطأ التلميحات التالية حول التدريب:

- يشير المنحدر الهابط إلى أن النموذج يتحسن.

- يشير المنحدر الصاعد إلى أن النموذج يزداد سوءًا.
- يشير المنحدر المسطح إلى أن النموذج قد وصل إلى نقطة التقارب (convergence).

على سبيل المثال، يوضح منحنى الخطأ المثالي إلى حد ما التالي:

- منحدر شديد الانحدار أثناء التكرارات الأولية، مما يعني تحسينًا سريعًا للنموذج.
- منحدر تدريجي (ولكن لا يزال هبوطًا) حتى يقترب من نهاية التدريب، مما يعني استمرار تحسين النموذج بوتيرة أبطأ إلى حد ما ثم خلال التكرارات الأولية.
- منحدر مسطح نحو نهاية التدريب مما يوحي بالتقارب.



على الرغم من أهمية خطأ التدريب، انظر أيضًا التعميم (generalization).

Training-serving skew

انحراف خدمة التدريب

الفرق بين أداء النموذج أثناء التدريب وأداء نفس النموذج أثناء الخدمة.

Training set

مجموعة التدريب

المجموعة الفرعية من مجموعة البيانات المستخدمة لتدريب نموذج.

تقليديا، يتم تقسيم الأمثلةفي مجموعة البيانات إلى المجموعات الفرعية الثلاثة المتميزة التالية:

- مجموعة التدريب (training set).
- مجموعة التحقق من الصحة (validation set).
- مجموعة الاختبار (test set).

من الناحية المثالية، يجب أن ينتمي كل مثال في مجموعة البيانات إلى واحدة فقط من المجموعات الفرعية السابقة. على سبيل المثال، يجب ألا ينتمي مثال واحد إلى كل من مجموعة التدريب ومجموعة التحقق من الصحة.

Trajectory

المسار

في التعلم المعزز (reinforcement learning)، سلسلة من الصفوف (tuples) التي تمثل سلسلة من انتقالات الحالة (state) للوكيل (agent)، حيث تتوافق كل مجموعة مع الحالة، والاجراء (action)، والمكافأة (reward)، والحالة التالية لانتقال حالة معينة.

Transfer learning

نقل التعلم

نقل المعلومات من مهمة تعلم الآلة إلى أخرى. على سبيل المثال، في التعلم متعدد المهام (multi-task learning)، يحل نموذج واحد مهامًا متعددة، مثل نموذج عميق (deep model) يحتوي على عقد إخراج مختلفة لمهام مختلفة. قد يتضمن التعلم عن طريق نقل المعرفة من حل مهمة أبسط إلى مهمة أكثر تعقيداً، أو ينطوي على نقل المعرفة من مهمة حيث يوجد المزيد من البيانات إلى واحد حيث توجد بيانات أقل.

تحل معظم أنظمة التعلم الآلي مهمة واحدة. نقل التعلم هو خطوة صغيرة نحو الذكاء الاصطناعي حيث يمكن لبرنامج واحد حل مهام متعددة.

Transformer

المحول

بنية شبكة عصبية (neural network) تم تطويرها في Google وتعتمد على آليات الانتباه الذاتي (self-attention) لتحويل تسلسل تضمين المدخلات إلى سلسلة من عمليات دمج المخرجات دون الاعتماد على التلافيف (convolutions) أو الشبكات العصبية المتكررة (recurrent neural networks). يمكن اعتبار المحول مكدس من طبقات الانتباه الذاتي.

يمكن أن يتضمن المحول أيًا مما يلي:

- المشفر (encoder).
- مفكك الشفرة (decoder).
- كلاً من المشفر ومفكك الشفرة.

يحول المشفر سلسلة من التضمينات (embeddings) إلى تسلسل جديد بنفس الطول. يشتمل المشفر على N طبقات متطابقة، تحتوي كل منها على طبقتين فرعيتين. يتم تطبيق هاتين الطبقتين الفرعيتين في كل موضع من تسلسل تضمين الإدخال، مما يحول كل عنصر من عناصر التسلسل إلى تضمين جديد. تجمع الطبقة الفرعية الأولى من المشفر المعلومات عبر تسلسل الإدخال. تقوم الطبقة الفرعية الثانية للمشفر بتحويل المعلومات المجمعة إلى دمج مخرجات.

يحول مفكك الشفرة سلسلة من إدخالات المدخلات إلى سلسلة من التضمينات المخرجة، ربما بطول مختلف. يشتمل مفكك الشفرة أيضًا على طبقات متطابقة N مع ثلاث طبقات فرعية، اثنتان منها تشبه الطبقات الفرعية للمشفر. تأخذ الطبقة الفرعية الثالثة من مفكك الشفرة إخراج المشفر وتطبق آلية الانتباه الذاتي لجمع المعلومات منه.

Translational invariance

ثبات الترجمة

في مشكلة تصنيف الصور، قدرة الخوارزمية على تصنيف الصور بنجاح حتى عندما يتغير موضع الكائنات داخل الصورة. على سبيل المثال، لا يزال بإمكان الخوارزمية تحديد كلب، سواء كان في وسط الإطار أو في الطرف الأيسر من الإطار.

انظر أيضًا ثبات الحجم (size invariance) وثبات الدوران (rotational invariance).

Trigram

تريجرام

تريجرام هو N جرام (N-gram) فيه N = 3.

True positive (TP)

صحيح سلبي (TN)

مثال يتنبأ فيه النموذج بشكل صحيح بالفئة السلبية (negative class). على سبيل المثال، يستنتج النموذج أن رسالة بريد إلكتروني معينة ليست بريدًا عشوائيًا، وأن رسالة البريد الإلكتروني هذه ليست في الحقيقة بريدًا عشوائيًا.

True positive (TP)

صحيح ايجابي (TP)

مثال يتنبأ فيه النموذج بشكل صحيح بالفئة الإيجابية (positive class). على سبيل المثال، يستنتج النموذج أن رسالة بريد إلكتروني معينة هي بريد عشوائي، وأن رسالة البريد الإلكتروني هذه هي بالفعل بريد عشوائي.

True positive rate (TPR)

مرادف للاسترجاع (recall). هذا هو:

$$\text{true positive rate} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

المعدل الحقيقي الإيجابي هو المحور yفي منحنى ROC (ROC curve).

Type 1 Error

خطأ من النوع الأول

ايجابيات خاطئة (False Positives). ضعفي اعتبارك شركة تقوم بتحسين ممارسات التوظيف لتقليل الإيجابيات الخاطئةفي عروض العمل. يحدث خطأ من النوع الأول عندما يبدو المرشح جيداً ويقومون بتوظيفه، لكنه سيء بالفعل.

Type 2 Error

خطأ من النوع الثاني

السلبيات الخاطئة (False Negatives). كان المرشح عظيما لكن الشركة فاتته.

U

U

Underfitting

الضبط الناقص

إنتاج نموذج بقدرة تنبؤية ضعيفة لأن النموذج لم يلتقط بشكل كامل مدى تعقيد بيانات التدريب. يمكن أن تسبب العديد من المشاكل الضبط الناقص، بمافي ذلك:

- التدريب على مجموعة خاطئة من الميزات (features).
- التدريب لفترات (epochs) قليلة جداً أو بمعدل تعلم (learning rate) منخفض جداً.
- التدريب بمعدل تنظيم (regularization rate) مرتفع للغاية.
- توفير عدد قليل جداً من الطبقات المخفية (hidden layers)في شبكة عصبية عميقة.

Undersampling

اختزال العينات

إزالة أمثلة (examples) من فئة الأغلبية (majority class)في مجموعة بيانات غير متوازنةفي الفئة (class-imbalanced dataset) من أجل إنشاء مجموعة تدريب (training set) أكثر توازناً.

على سبيل المثال، ضعفي اعتبارك مجموعة بيانات تكون فيها نسبة فئة الأغلبية (majority class) إلى فئة الأقلية (minority class) 20: 1. للتغلب على هذا الخللفي التوازنفي الفئة، يمكنك إنشاء مجموعة تدريب تتكون من جميع أمثلة فئات الأقلية ولكن فقط عُشر أمثلة فئة الأغلبية، والتي من شأنها إنشاء نسبة فئة مجموعة التدريب 2: 1. بفضل اختزال العينات، قد تنتج مجموعة التدريب الأكثر توازناً هذه نموذجًا أفضل. بدلاً من ذلك، قد تحتوي مجموعة التدريب الأكثر توازناً على أمثلة غير كافية لتدريب نموذج فعال.

على النقيض من الإفراطفي أخذ العينات (oversampling).

U-net

تم اختراع U-netفي الأصل واستخدم لأول مرة لتجزئة الصور الطبية الحيوية (biomedical image segmentation). يمكن اعتبار بُنيتها على نطاق واسع على أنها شبكة مشفر (encoder) متبوعة بشبكة مفكك شفرة (decoder).

Unidirectional

أحادي الاتجاه

نظام يقوم فقط بتقييم النص الذي يسبق القسم الهدف من النص.في المقابل، يقوم النظام ثنائي الاتجاه بتقييم كل من النص الذي يسبق ويتبع القسم المستهدف من النص.

انظر ثنائية الاتجاه (bidirectional) لمزيد من التفاصيل.

Unidirectional language model

نموذج لغة أحادي الاتجاه

نموذج اللغة (language model) الذي يبني احتمالاته فقط على الرموز المميزة (token) التي تظهر قبل، وليس بعد، الرمز (الرموز) الهدف. على النقيض من نموذج اللغة ثنائي الاتجاه (bidirectional language model).

Unlabeled example

مثال غير مسمى (غير مصنف)

مثال يحتوي على ميزات (features) ولكن بدون تسمية (label). على سبيل المثال، يوضح الجدول التالي ثلاثة أمثلة غير مسماة من نموذج تقييم المنزل، ولكل منها ثلاث ميزات ولكن ليس لها قيمة منزلية:

Number of bedrooms	Number of bathrooms	House age
3	2	15
2	1	72
4	2	34

في التعلم الآلي الخاضع للإشراف (supervised machine learning)، تتدرب النماذج على الأمثلة المصنفة وتتنبأ بالأمثلة غير المسماة (unlabeled examples).

في التعلم شبه الخاضع للإشراف (semi-supervised) وغير الخاضع للإشراف (unsupervised)، يتم استخدام أمثلة غير مسماة أثناء التدريب.

يتناقض المثال غير المسمى مع المثال المسمى (labeled example).

Unsupervised machine learning

التعلم الآلي غير الخاضع للإشراف

تدريب نموذج (model) للعثور على أنماطفي مجموعة بيانات، عادةً مجموعة بيانات غير مسماة.

الاستخدام الأكثر شيوعًا للتعلم الآلي غير الخاضع للإشراف هو تجميع (cluster) البياناتفي مجموعات من الأمثلة المتشابهة. على سبيل المثال، يمكن لخوارزمية التعلم الآلي غير الخاضعة للاشراف أن تجمع الأغاني بناءً على الخصائص المختلفة للموسيقى. يمكن أن تصبح المجموعات الناتجة مدخلاً لخوارزميات تعلم الآلة الأخرى (على سبيل المثال، لخدمة توصية الموسيقى). يمكن أن يساعد التجميع عندما تكون التسميات المفيدة نادرة أو غائبة. على سبيل المثال،في مجالات مثل مكافحة الإساءة (anti-abuse) والاحتيال (fraud)، يمكن أن تساعد المجموعات البشرية على فهم البيانات بشكل أفضل.

على النقيض من التعلم الآلي الخاضع للإشراف (supervised machine learning).

Uplift modeling

النمذجة المرتفعة

أسلوب نمذجة، يشيع استخدامهفي التسويق، يضع نموذجًا "للتأثير السببي "causal effect (المعروف أيضًا باسم "التأثير المتزايد "incremental impact") "للعلاج "treatment على "الفرد "individual". فيما يلي مثالان:

- قد يستخدم الأطباء نمذجة الارتقاء للتنبؤ بانخفاض معدل الوفيات (التأثير السببي) لإجراء طبي (علاج) اعتمادًا على العمر والتاريخ الطبي للمريض (الفرد).
- قد يستخدم المسوقون النمذجة المرتفعة للتنبؤ بزيادة احتمالية الشراء (التأثير السببي) بسبب إعلان (علاج) على شخص (فرد).

تختلف النمذجة المرتفعة عن التصنيف أو الانحدارفي أن بعض التسميات (على سبيل المثال، نصف السمياتفي المعالجات الثنائية) مفقودة دائمًافي النمذجة المرتفعة. على سبيل المثال، يمكن للمريض أن يتلقى أو لا يتلقى العلاج؛ لذلك، يمكننا فقط ملاحظة ما إذا كان المريض سيشفى أم لافي حالة واحدة فقط من هاتين الحالتين (ولكن ليس كلاهما). الميزة الرئيسية لنموذج الارتقاء (uplift model) هي أنه يمكن أن يولد تنبؤات للوضع غير المرصود (الواقع المضاد) ويستخدمه لحساب التأثير السببي.

upweighting

زيادة الوزن

تطبيق وزن على فئة تم اختزال (downsampled) العينة يساوي العامل الذي اختزلت من خلاله.

User matrix

مصفوفة المستخدم

في أنظمة التوصية (recommendation systems)، متجه التضمين (embedding vector) الناتج عن عامل المصفوفة (matrix factorization) الذي يحمل إشارات كامنة حول تفضيلات المستخدم. يحتوي كل صف من مصفوفة المستخدم على معلومات حول القوة النسبية للإشارات الكامنة المختلفة لمستخدم واحد. على سبيل المثال، ضعفي اعتبارك نظام توصية للأفلام.في هذا النظام، قد تمثل الإشارات الكامنةفي مصفوفة المستخدم اهتمام كل مستخدم بأنواع معينة، أو قد يكون من الصعب تفسير الإشارات التي تتضمن تفاعلات معقدة عبر عوامل متعددة.

تحتوي مصفوفة المستخدم على عمود لكل ميزة كامنة وصف لكل مستخدم. أي أن مصفوفة المستخدم لها نفس عدد الصفوف مثل المصفوفة المستهدفة التي يتم تحليلها. على سبيل المثال، بالنظر إلى نظام توصية الأفلام لمليون مستخدم، فإن مصفوفة المستخدم ستحتوي على 1000000 صف.

V

V

Validation

التحقق من الصحة

التقييم المبدئي لجودة النموذج. يتحقق التحقق من الصحة من جودة تنبؤات النموذج مقابل مجموعة التحقق من الصحة (validation set).

نظرًا لاختلاف مجموعة التحقق من الصحة عن مجموعة التدريب (training set)، فإن التحقق من الصحة يساعد في الحماية من الضبط الزائد (overfitting).

قد تفكر في تقييم النموذج مقابل مجموعة التحقق من الصحة كالجولة الأولى من اختبار وتقييم النموذج مقابل مجموعة الاختبار (test set) كجولة ثانية من الاختبار.

Validation loss

خطأ التحقق من الصحة

مقياس (metric) يمثل خطأ (loss) النموذج في مجموعة التحقق من الصحة (validation set) أثناء تكرار (iteration) معين للتدريب.

انظر أيضا منحنى التعميم (generalization curve).

Validation set

مجموعة التحقق من الصحة

المجموعة الفرعية من مجموعة البيانات (dataset) التي تجري تقييمًا أوليًا مقابل نموذج (model) مدرب. عادةً ما تقوم بتقييم النموذج المدرب مقابل مجموعة التحقق من الصحة (validation set) عدة مرات قبل تقييم النموذج مقابل مجموعة الاختبار (test set).

تقليديا، تقسم الأمثلة في مجموعة البيانات إلى المجموعات الفرعية الثلاثة المتميزة التالية:

- مجموعة التدريب (training set).
- مجموعة التحقق من الصحة (validation set).
- مجموعة الاختبار (test set).

من الناحية المثالية، يجب أن ينتمي كل مثال في مجموعة البيانات إلى واحدة فقط من المجموعات الفرعية السابقة. على سبيل المثال، يجب ألا ينتمي مثال واحد إلى كل من مجموعة التدريب ومجموعة التحقق من الصحة.

Vanishing gradient problem

مشكلة تلاشي التدرج

الميل إلى أن تصبح تدرجات الطبقات المخفية (hidden layers) المبكرة لبعض الشبكات العصبية العميقة (deep neural networks) مسطحة بشكل مدهش (منخفضة). تؤدي التدرجات المنخفضة بشكل متزايد إلى تغييرات أصغر بشكل متزايدفي الأوزان الموجودة على العقدفي الشبكة العصبية العميقة، مما يؤدي إلى القليل من التعلم أو انعدامه. تصبح النماذج التي تعاني من مشكلة التدرج المتلاشي صعبة أو مستحيلة التدريب. تعالج خلايا الذاكرة طويلة قصيرة المدى (LSTM) هذه المشكلة.

قارن مع مشكلة انفجار التدرج (exploding gradient problem).

variable importance

أهمية المتغير

مجموعة من الدرجات التي تشير إلى الأهمية النسبية لكل ميزة (feature) للنموذج.

على سبيل المثال، ضعفي اعتبارك شجرة قرار (decision tree) تقدر أسعار المنازل. افترض أن شجرة القرار هذه تستخدم ثلاث ميزات: الحجم والعمر والنمط. إذا تم حساب مجموعة من الواردات المتغيرة للسمات الثلاث على أنها {الحجم = 5.8 ، العمر = 2.5 ، النمط = 4.7} ، فإن الحجم يكون أكثر أهمية لشجرة القرار من العمر أو النمط.

توجد مقاييس مختلفة متغيرة الأهمية، والتي يمكن أن تطلع خبراء تعلم الآلة على جوانب مختلفة من النماذج.

Variance

التباين

ما مدى تباين توقعاتك لملاحظة معينة بالنسبة لبعضها البعض؟

- يشير التباين المنخفض (Low variance) إلى أن نموذجك متسق داخليًا، مع اختلاف التوقعات قليلاً عن بعضها البعض بعد كل تكرار.
- يشير التباين الكبير (High variance) (مع انحياز منخفض (low bias)) إلى أن نموذجك قد يكون أكثر من اللازم ويقرأ بعمق شديدفي الضوضاء الموجودةفي كل مجموعة تدريب.

VGG Network

VGG اختصارا لـ Visual Geometry Group ؛ إنها بنية شبكة عصبية تلافيفية عميقة قياسية (CNN) ذات طبقات متعددة. يشير "العمق" إلى عدد الطبقات ذات VGG-16 أو VGG-19 المكونة من 16 و 19 طبقة تلافيفية (convolutional layer). تعد بُنية VGG هي الأساس لنماذج التعرف على الأشياء (object recognition).





Weight

الوزن

قيمة يضربها النموذجفي قيمة أخرى. التدريب (Training) هو عملية تحديد أوزان النموذج المثالية؛ الاستدلال (inference) هو عملية استخدام تلك الأوزان المكتسبة لعمل تنبؤات.

Weighted sum

المجموع المرجح

مجموع كل قيم الإدخال ذات الصلة مضروبةفي الأوزان المقابلة لها. على سبيل المثال، افترض أن المدخلات ذات الصلة تتكون مما يلي:

input value	input weight
2	-1.3
-1	0.6
3	0.4

The weighted sum is therefore:

وبالتالي فإن المجموع المرجح هو:

weighted sum = (2)(-1.3) + (-1)(0.6) + (3)(0.4) = -2.0

المجموع المرجح هو وسيطة الإدخال لدالة التنشيط (activation function).

Wide model

نموذج واسع

نموذج خطي يحتوي عادةً على العديد من ميزات الإدخال المتفرقة (sparse input features). نشير إليها على أنها "واسعة" لأن مثل هذا النموذج هو نوع خاص من الشبكات العصبية (neural network) مع عدد كبير من المدخلات التي تتصل مباشرة بعقدة الإخراج. غالبًا ما تكون النماذج العريضة أسهلفي التصحيح والفحص من النماذج العميقة (deep models). على الرغم من أن النماذج الواسعة لا يمكنها التعبير عن اللاخطية من خلال الطبقات المخفية (hidden layers)، إلا أن النماذج الواسعة يمكن أن تستخدم تحويلات مثل تقاطع الميزات (feature crossing) والتجميع (bucketization) لنمذجة اللاخطية بطرق مختلفة.

على النقيض من النموذج العميق (deep model).

Width

العرض

عدد الخلايا العصبية (neuron)في طبقة (layer) معينة من الشبكة العصبية (neural network).

wisdom of the crowd

حكمة الجموع

غالبًا ما تؤدي الفكرة القائلة بأن متوسط آراء أو تقديرات مجموعة كبيرة من الناس ("الجمهور") إلى نتائج جيدة بشكل مدهش. على سبيل المثال، ضعفي اعتبارك لعبة يخمن فيها الناس عدد حبوب الهلام المعبأةفي وعاء كبير. على الرغم من أن معظم التخمينات الفردية ستكون غير دقيقة، فقد ثبت بشكل تجريبي أن متوسط جميع التخمينات قريب بشكل مدهش من العدد الفعلي لحبوب الهلامفي الجرة.

المجاميع (Ensembles) هي نظير برمجي لحكمة الجموع. حتى لو قدمت النماذج الفردية تنبؤات غير دقيقة إلى حد كبير، فإن حساب متوسط تنبؤات العديد من النماذج غالبًا ما يولد تنبؤات جيدة بشكل مدهش. على سبيل المثال، على الرغم من أن شجرة القرار (decision tree) الفردية قد تقدم تنبؤات سيئة، إلا أن مجموعة القرارات (decision forest) غالبًا ما تقدم تنبؤات جيدة جداً.

Word embedding

تضمين الكلمة

تمثيل (Representing) كل كلمةفي مجموعة كلمات داخل متجه تضمين (embedding vector)؛ أي تمثيل كل كلمة كمتجه لقيم الفاصلة العائمة بين 0.0 و 1.0. الكلمات ذات المعاني المتشابهة لها تمثيلات أكثر تشابهًا من الكلمات ذات المعاني المختلفة. على سبيل المثال، سيكون للجزر والكرفس والخيار تمثيلات متشابهة نسبيًا، والتي ستكون مختلفة تمامًا عن تمثيلات الطائرات والنظارات الشمسية ومعجون الأسنان.

Word2vec

Word2vec هي تقنية لمعالجة اللغة الطبيعية (NLP) نُشرتفي عام 2013. تستخدم خوارزمية word2vec نموذج شبكة عصبية لتعلم ارتباطات الكلمات من مجموعة كبيرة من النصوص. بمجرد التدريب، يمكن لهذا النموذج اكتشاف الكلمات المترادفة أو اقتراح كلمات إضافية لجملة جزئية.

المصادر

1. Data Science Glossary, https://www.datascienceblog.net/data-science-glossary/
2. Deep Learning Glossary, https://deeplearning.lipingyang.org/deep-learning-glossary/
3. Deep Learning Glossary: Machine Learning Terms Definitions &Acronyms, https://www.kiranvoleti.com/deep-learning-glossary
4. Glossary of common Machine Learning, Statistics and Data Science terms, https://www.analyticsvidhya.com/glossary-of-common-statistics-and-machine-learning-terms/
5. Glossary of common Machine Learning, Statistics and Data Science terms, https://www.analyticsvidhya.com/glossary-of-common-statistics-and-machine-learning-terms/
6. Glossary of Terms, https://ai.stanford.edu/~ronnyk/glossary.html
7. Machine Learning Glossary, https://developers.google.com/machine-learning/glossary?hl=en
8. Machine Learning Glossary, https://ml-cheatsheet.readthedocs.io/en/latest/glossary.html
9. Machine Learning Glossary, https://thecleverprogrammer.com/2021/11/18/machine-learning-glossary/
10. The Most Common Machine Learning Terms, Explained, https://www.springboard.com/blog/data-science/machine-learning-terminology/

Machine Learning Deep Learning Data Science Glossary

By: Dr. Alaa Taima

